

Waves of Collapse

Compression as the Geometry of Reachable Description

A Monograph on Hierarchical Templates, Continual Rematching,
and the Convergence of Archives into World Models

Flyxion

June 2026

Preface

This monograph began with a deceptively simple engineering idea: an audio compressor that builds a hierarchy of templates and continually rematches its archive against them, so that compression improves *in waves* as the system accumulates experience. The idea is easy to state and, in a narrow sense, easy to prototype. But pulled at, the thread unravels the entire fabric of what compression *is*.

A codec that rewrites its past whenever it learns a new regularity is not a codec in the classical sense. It is an epistemic agent. Its archive is not a warehouse of frozen bitstreams but a living theory of the data it has seen, and each newly discovered template is a small scientific revolution: a reinterpretation of old evidence under a richer ontology. Compression ratio becomes a measure of understanding; the punctuated “collapse events” in which large regions of the archive suddenly shrink are precisely the ontology revisions studied in *Persistent Anomalies and the Geometry of Ontology Revision*; and the template hierarchy is an admissibility structure in the sense of the Frozen Processes program — a stratified space of reachable descriptions in which the primitive notion is not *what a signal is* but *what encodings of it can be reached from the current library*.

The monograph therefore has two registers. One is concrete: data structures, rematching schedules, grammar inference, rate–distortion trade-offs, multimodal fusion, and the engineering of an archive that treats every photograph, LiDAR sweep, and microphone recording as another slice through a shared latent world. The other is foundational: the claim that compression, prediction, inference, memory, and scientific theory change are a single phenomenon — relaxation in a constraint field — and that the RSVP formalism of scalar density Φ , vector flow \mathbf{v} , and entropy S gives this phenomenon a field-theoretic home.

Readers who want the algorithm may read Parts I and II and stop. Readers who want the ontology may begin at Part IV. The author’s hope is that most will discover, as he did, that the two cannot finally be separated: the engineering forces the philosophy, and the philosophy repays the debt by predicting the engineering’s failure modes.

Flyxion
June 2026

Contents

I	The Problem of the Frozen Codec	1
1	Compression Without Memory	2
1.1	The classical setting	2
1.2	The proposal in one page	2
1.3	Why this is not merely a better dictionary coder	3
2	A Formal Skeleton	5
2.1	Templates, encodings, and the description complex	5
2.2	Rematching as relaxation	5
2.3	Collapse events and the wave structure	7
2.4	The description complex as a sheaf	8
II	Mechanics of a Rematching Archive	9
3	The Template Hierarchy in Practice	10
3.1	Level 1: fragments and the quantization floor	10
3.2	Level 2 and above: grammar over the fragment alphabet	10
3.3	Rematching schedules	10
3.4	Lossy regimes and the distortion lattice	11
4	Multimodal Fusion: From Codec to World Model	12
4.1	Observations as slices	12
4.2	A worked cascade: when sound rewrites a photograph	12
4.3	The crossover point	13
4.4	Residuals as the archive’s frontier	15
III	Information-Theoretic Foundations	16
5	Two-Part Codes, MDL, and the Rent a Template Pays	17
5.1	Kolmogorov ideal and computable practice	17
5.2	MDL admission as Bayesian model comparison	17
5.3	Compression = prediction = comprehension	18

<i>CONTENTS</i>	iii
5.4 The Weierstrass lesson: roughness, typicality, and forced strata	18
6 The Thermodynamics of Rematching	20
6.1 Description length as free energy	20
6.2 RSVP: the field-theoretic reading	20
6.3 TARTAN and the texture of the residual field	21
IV Admissibility, Strata, and Frozen Processes	23
7 The Archive as a Stratified Reachability Space	24
7.1 Strata of description	24
7.2 Persistent anomalies and forced revision	24
7.3 Belief geometry of the encoder	25
8 Memory, Brains, and Resonance Without a Substrate	27
9 Threads, Projections, and the Category of Admissible Paths	29
9.1 Against modality-first typing	29
9.2 Yarncrawler threads as the primitive type	29
9.3 Geometric and algebraic at once	30
9.4 The loop foundation models do not close	30
10 Compression as Epistemology	31
10.1 The archive as a scientific community	31
10.2 Knowledge as navigation	31
V Consequences, Designs, and Open Problems	32
11 Engineering Consequences	33
11.1 What changes when size is not a property of files	33
11.2 Failure modes the theory predicts	33
11.3 A minimal experimental program	34
12 Open Problems	36
Coda	37
Bibliographic Notes	38

Part I

The Problem of the Frozen Codec

Chapter 1

Compression Without Memory

1.1 The classical setting

Classical lossless compression is a contract between two parties who share a prior. Shannon's source coding theorem fixes the terms: for a stationary ergodic source X with entropy rate $H(X)$, no uniquely decodable code can achieve an expected rate below $H(X)$ bits per symbol, and arithmetic coding approaches it given the true source distribution. Every practical codec is an approximation of this contract under ignorance of the true distribution: it must either assume a model (transform codecs assume local spectral sparsity and psychoacoustic masking; predictive codecs like FLAC assume short-range linear predictability) or learn one on the fly within a bounded window (LZ77's sliding dictionary, adaptive arithmetic coders, context mixing).

Three structural commitments are shared by essentially all deployed codecs, and all three are targets of this monograph.

1. **Fixed representational basis.** The vocabulary in which the signal is described — MDCT coefficients, LPC residuals, dictionary phrases — is chosen at design time or adapted only within a short horizon. The codec cannot discover that drum hits exist.
2. **Write-once encoding.** Once a block is compressed, its encoding is final. Nothing learned later improves it. The archive is a graveyard of past states of knowledge.
3. **Modality isolation.** Audio codecs know nothing of video; image codecs know nothing of geometry. Each modality is compressed against a model that cannot be informed by any other.

These commitments are not errors. They purchase locality, streaming, bounded memory, and decoder simplicity — the virtues required when a codec is a thin pipe between sender and receiver. But an *archive* is not a pipe. An archive persists, accretes, and is consulted across years. For an archive, the three commitments are pathologies: they freeze the system's ontology at the moment of ingestion.

1.2 The proposal in one page

The system studied here, which we will call a *rematching archive*, inverts all three commitments.

Definition 1.1 (Rematching archive, informal). *A rematching archive maintains (i) a growing hierarchy of templates $\mathcal{T} = \mathcal{T}_1 \cup \mathcal{T}_2 \cup \dots$, where level- k templates are compositions of level- $(k-1)$ templates; (ii) an encoding of every stored observation as a path through template space plus a residual; and (iii) a background process that, whenever the hierarchy changes, re-evaluates existing encodings and rewrites any that the new hierarchy explains more cheaply.*

At the lowest level, templates are short waveform fragments, image patches, or point-cloud surfels:

$$\mathcal{T}_1 = \{t_1, t_2, \dots, t_n\}.$$

The encoder initially stores incoming data nearly literally, because the library explains nothing. As recurrence appears, segments are replaced by references. Higher levels compose lower ones,

$$\mathcal{T}_2 \ni \tau = (t_3, t_7, t_2), \quad \mathcal{T}_3 \ni T = (\tau^{(5)}, \tau^{(11)}, \tau^{(2)}),$$

until a drum hit is a single reference, a fill is a sequence of hits, a chorus is a sequence of fills, and a song is a short program over the library.

The decisive feature is *retroactivity*. Suppose the archive holds the encoded stream $A B C D E$ and later discovers the template $X = A B C$. The stream is rewritten as $X D E$. Later still $Y = X D$ appears and the stream collapses to $Y E$. Compression improves after the fact, and not smoothly: each genuinely new template triggers a cascade of rewrites whose aggregate effect is a sudden drop in archive size — a *collapse event*. Empirically one expects a compression trajectory like

$$1.0\times \rightarrow 1.3\times \rightarrow 1.8\times \rightarrow 2.7\times \rightarrow 4.1\times \rightarrow \dots$$

with the jumps clustered around discoveries rather than spread uniformly over time. We will call this trajectory, recorded jump by jump in the archive’s event ledger, the *Admissibility Log trail*: the append-only record of which descriptions became reachable, and when. Compression occurs in waves because understanding occurs in waves.

1.3 Why this is not merely a better dictionary coder

Grammar-based compressors (Sequitur, Re-Pair), vector quantization, sparse coding, and fractal compression each contain a fragment of the idea. Sequitur builds hierarchical rules online; Re-Pair greedily replaces frequent pairs; VQ learns a codebook; sparse coding learns an overcomplete basis; fractal methods exploit self-similarity across scales. The rematching archive differs on three axes simultaneously:

1. the library grows *without bound in principle*, subject only to a minimum-description-length (MDL) accounting that charges every template its own storage cost;
2. old encodings are *repeatedly* re-evaluated, so the representation of a fixed object is a function of the whole archive’s history, not of the object alone;
3. the library is *shared across modalities*, so that a LiDAR scan can reduce the encoded size of a photograph taken eight years earlier.

The third axis is where the codec framing finally breaks. Once templates are allowed to be *generative models of latent structure* — a chair’s geometry, a room’s acoustics, a building’s facade — rather than literal signal fragments, the archive stops storing observations and starts storing a world, with observations demoted to evidence plus residuals. Part III develops this transition; Parts IV and V argue that it was implicit from the first line of code.

The opposition can be tabulated, and the table is worth keeping in view throughout, because every later chapter elaborates one of its three rows.

	The frozen codec	The rematching archive
Basis	Fixed at design time	Grows without bound via the hierarchical template library
Encoding	Write-once and static	Retroactive; continually rematched against the library
Scope	Modality-isolated	Shared across modalities as latent world structure

Table 1.1: The three structural commitments of classical compression and their inversions. An archive is not a pipe; it is an epistemic agent, and its compression ratio is a direct, if stratum-relative, measure of its understanding.

Chapter 2

A Formal Skeleton

2.1 Templates, encodings, and the description complex

Fix a space of raw observations \mathcal{O} (finite strings over sample alphabets, in practice). A *template library* is a finite set \mathcal{T} together with a realization map $\rho : \mathcal{T} \rightarrow \text{Programs}$ assigning each template a decoding program. Level structure is induced, not imposed: $\text{level}(t) = 1$ if $\rho(t)$ references no other template, and $\text{level}(t) = 1 + \max\{\text{level}(s) : s \text{ referenced by } t\}$ otherwise. The library is thus a DAG; we call its order complex the *description complex* of the archive.

Definition 2.1 (Encoding). *An encoding of $o \in \mathcal{O}$ over \mathcal{T} is a pair (π, r) where π is a finite sequence of references into \mathcal{T} (with arguments) and r is a residual string, such that the decoder reconstructs o exactly (lossless regime) or within distortion δ (lossy regime) from (π, r) . Its cost is $L(o | \mathcal{T}) = |\pi| + |r|$ in bits under the archive's entropy coder.*

Definition 2.2 (Archive state). *An archive state is a triple $(\mathcal{T}, E, \Lambda)$ where E assigns each stored observation an encoding over \mathcal{T} and $\Lambda = L(\mathcal{T}) + \sum_o L(o | \mathcal{T})$ is the total two-part description length, charging the library its own storage.*

The two-part form is essential. Without charging $L(\mathcal{T})$, the system degenerates: it can always add a template equal to any observation and “compress” it to one reference. MDL accounting makes template admission a genuine hypothesis test: a template is worth adding exactly when the total Λ decreases, i.e. when the regularity it names is real enough to pay its own rent.

2.2 Rematching as relaxation

Let \mathcal{S} be the set of archive states over a fixed multiset of stored observations. Define a directed graph on \mathcal{S} whose edges are the *admissible moves*:

- **ADMIT**(t): add template t to \mathcal{T} ;
- **RETIRE**(t): remove an unreferenced template;
- **REWRITE**(o): replace $E(o)$ with a cheaper encoding over the current \mathcal{T} ;

- $\text{REFACTOR}(t, t')$: rewrite the realization of t in terms of t' (template-to-template compression).

Definition 2.3 (Reachable optimum). $\Lambda^*(\sigma) = \min\{\Lambda(\sigma') : \sigma' \in \text{Reach}(\sigma)\}$, the least description length reachable from state σ by admissible moves.

This is the monograph's first load-bearing wall. The archive never optimizes over all mathematically possible representations — that target is the (uncomputable) Kolmogorov complexity K of the corpus. It optimizes over representations *reachable from where it stands*, through moves it can actually perform, at costs it can actually pay. The gap between $\Lambda^*(\sigma)$ and $K(\text{corpus})$ is not an implementation defect; it is the constitutive structure of the system, in exactly the sense in which the admissibility/reachability program takes reachability rather than existence as the foundational primitive. A representation that exists but is not reachable from the current library might as well not exist; a representation becomes available only when the hierarchy has grown a path to it.

The gap admits a sharper statement. Let $\mathcal{A}(\sigma)$ denote the set of admissible moves from σ , and let a *constraint path* be a sequence $\gamma = (\sigma_0, \sigma_1, \dots, \sigma_n)$ with $\sigma_{i+1} \in \mathcal{A}(\sigma_i)$. Then

$$\Lambda^*(\sigma_0) = \min_{\gamma: \gamma(0)=\sigma_0} \Lambda(\gamma(n)) = K(D) + \delta_{\mathcal{T}}(\sigma_0),$$

where $\delta_{\mathcal{T}}(\sigma_0) \geq 0$ is the *ontological deficit*: the frozen invariant of the current stratum, the portion of description length that cannot be relaxed from σ_0 by any admissible path without an ADMIT that breaks the stratum's boundary conditions. The deficit is not an error term to be apologized for. It is the quantitative measure of what the archive's present ontology cannot say, and its decrease over time — stepwise, at collapse events — is the archive's intellectual history. The Kolmogorov limit is recovered only in the idealized closure $\delta_{\mathcal{T}} \rightarrow 0$, which no finite history attains.

Proposition 2.4 (Monotone relaxation). *Under any schedule that performs only moves with $\Delta\Lambda \leq 0$, the sequence $\Lambda(\sigma_0), \Lambda(\sigma_1), \dots$ is nonincreasing and converges. If ties are broken consistently and the residual coder is fixed, the process reaches a local minimum of Λ over $\text{Reach}(\sigma_0)$ in finitely many strict moves.*

Proof sketch. Λ is bounded below by 0 and each strict move decreases it by at least one bit; nonstrict moves can be bounded by canonical-form tie-breaking. Hence no infinite strictly decreasing sequence exists and the relaxation halts at a state from which no admissible move improves Λ . \square

The interesting dynamics live in the *non-monotone* regime: admitting a template momentarily *increases* Λ (the library got bigger; nothing has been rewritten yet) and pays off only after the rematching cascade propagates. The archive must therefore tolerate transient uphill moves — it must speculate. Section 4.3 treats admission as hypothesis-posting with deferred settlement, and Part IV identifies this tolerance for transient cost with the energy barrier between strata in the frozen-process picture.

2.3 Collapse events and the wave structure

Definition 2.5 (Collapse event). *A collapse event of magnitude m is a maximal cascade of REWRITE and REFACTOR moves triggered by a single ADMIT, with total description-length decrease m .*

Why waves rather than steady improvement? Because the payoff of a template is super-linear in its generality. A template explaining one segment saves a few bits; a template that turns out to be *the drum hit of the entire album* rewrites thousands of segments at once, and — crucially — its admission changes the statistics over which the entropy coder operates, cheapening every reference to it and re-ranking which *further* templates are worth admitting. Discovery is autocatalytic: each collapse flattens the landscape around it and exposes the next ridge. The size distribution of collapse events is conjectured (and in toy experiments observed) to be heavy-tailed:

Conjecture 2.6 (Punctuated compression). *For natural data streams with hierarchical generative structure, the magnitudes of collapse events under MDL-greedy rematching follow an approximately power-law distribution, and the compression-ratio trajectory is a devil’s-staircase-like function of data ingested: long plateaus punctuated by jumps at template discoveries.*

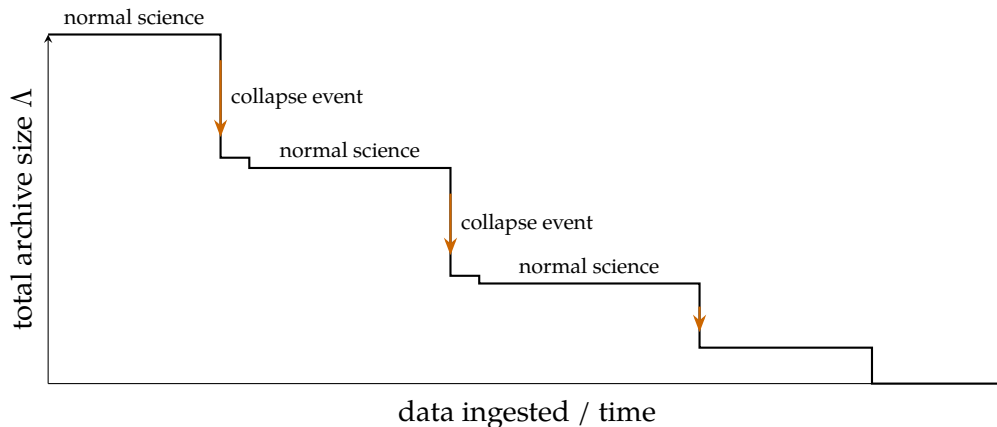


Figure 2.1: The Admissibility Log trail as a devil’s staircase. Plateaus are life within a stratum — routine encoding under the reigning ontology. Each vertical drop is a cascade triggered by a single admission: discovery flattens the landscape, rewrites thousands of old segments at once, and exposes the next ridge of structure.

The staircase is the empirical signature of the central claim: *the archive learns the way theories change*. Plateaus are normal science — routine encoding under the reigning ontology. Jumps are revisions — a new entity is admitted into the ontology and the past is reread in its terms. This is not a metaphor pasted on afterward; it is the same mathematics, and Chapter 7 makes the identification with the *Persistent Anomalies* account of ontology revision exact: a residual that stays stubbornly large under all reachable encodings is precisely a persistent anomaly, and the geometry of which template admissions can dissolve it is the geometry of ontology revision.

2.4 The description complex as a sheaf

The DAG structure of the library supports a finer reading. Regard the description complex as a base space whose open sets are families of mutually substitutable encodings, and assign to each open set the collection of valid local rewrites available over it. This assignment is a presheaf, and within a stratum it is a sheaf: local rewrites that agree on overlaps glue to a global rewrite, which is exactly why within-stratum relaxation is cheap, parallelizable, and order-insensitive. A collapse event is the failure case made productive: a local section — a rewrite valid over some region of the archive — that *fails to extend globally* under the current gluing data. Resolving the obstruction requires a REFACTOR that alters the base space itself: a topology-changing operation on the reference DAG, after which the section extends and the cascade propagates. Ontology revision, in this language, is forced gluing: the admission of a template introduces new overlap data under which previously incompatible local descriptions become sections of a single global one. The staircase is the trace of these gluings, and the persistent anomalies of Chapter 7 are the cohomological obstructions that no within-stratum section can absorb.

Part II

Mechanics of a Rematching Archive

Chapter 3

The Template Hierarchy in Practice

3.1 Level 1: fragments and the quantization floor

Level-1 audio templates are short fragments (2–50 ms) stored as quantized waveforms or, better, as points in a learned latent space with a fixed decoder. Matching is approximate: a segment matches t if the residual after subtracting (a time-warped, gain-scaled instance of) t codes more cheaply than the segment itself. The matcher’s invariance group — time shift, gain, mild warp, pitch transposition — is itself part of the model class, and enlarging it is one of the system’s discovery moves: learning that two fragments are *the same event at different gains* is learning a symmetry, and every learned symmetry is a quotient of the description space that shrinks the archive.

3.2 Level 2 and above: grammar over the fragment alphabet

Once level-1 references replace raw samples, the encoded stream is a string over a discrete alphabet, and the machinery of grammar-based compression applies directly: digram statistics, Re-Pair-style pair replacement, Sequitur-style online rule induction, with MDL admission replacing fixed heuristics. Level- k templates are phrases over level- $(k-1)$ symbols. Musical structure emerges without being asked for: hits compose into fills, fills into phrases, phrases into sections, because that is the actual generative grammar of the data and MDL is a consistent estimator of grammar under broad conditions.

3.3 Rematching schedules

Full rematching after every admission is quadratic and unnecessary. The practical design maintains an inverted index from templates to the encodings that reference them and from feature sketches (locality-sensitive hashes of segments) to candidate match sites. An admission enqueues only the encodings whose sketches collide with the new template; the cascade is processed with a priority queue ordered by expected savings. The archive thus has a foreground (ingest, encode greedily under current \mathcal{T}) and a background (relax toward Λ^*), and the background process is interruptible: the archive is at all times a valid, decodable, merely *suboptimal* representation. Identity is preserved across rewrites by an event-sourced

ledger — the Spherepop discipline — and the Spherepop vernacular makes the mechanism exact. An observation is not merely an ingestion event; it is a *nested semantic containment bubble*, a scope whose boundary conditions (what must decode to what, at what fidelity) are fixed at ingestion. A REWRITE is an evaluation that alters the internal mapping of the scope without violating its boundary: the bubble’s contents are repriced, never its identity. The “history of irreversible acts” is the append-only ledger of bubble evaluations, and “the same recording” has a stable referent across decades of rewrites because the referent was always the bubble, never the bits. Irreversibility lives in the ledger; reversibility lives in the representation; the boundary between them is enforced by scope nesting rather than by convention.

3.4 Lossy regimes and the distortion lattice

Everything above has a lossy counterpart in which the residual is coded to distortion δ under a perceptual metric. The novelty is that δ need not be fixed at ingestion: because the original-fidelity residual is retained until storage pressure forces a decision, the archive can *re-decide* its rate–distortion operating point per object as the library improves — old recordings can become *both* smaller and more faithful, because a better model spends its bits on genuine signal rather than on re-deriving structure it failed to recognize. Rate–distortion is not a curve the archive sits on; it is a curve the archive *moves*.

Chapter 4

Multimodal Fusion: From Codec to World Model

4.1 Observations as slices

Audio, video, still images, LiDAR, radar, text, and sensor logs are not distinct data types. They are projections of a shared world through different measurement kernels. A camera sees a chair from the front; a traditional codec stores pixels; the rematching archive initially does roughly the same, because it knows nothing. A second view from the side arrives, and the system can discover that two different images are projections of one 3D structure: the deeper template is the geometry, and both images become references to it plus view parameters plus residuals. A LiDAR sweep sharpens the geometry; every image of the object gets cheaper retroactively. A recording of the chair being dragged across the floor constrains its material; the material model sharpens rendering predictions; the images get cheaper again. Each modality is another slice through the latent structure

$$\mathcal{M} \rightarrow \mathcal{O}_i,$$

and as the latent model \mathcal{M} improves, every stored observation becomes more compressible. Observations collected years apart compress each other. A 2018 photograph helps compress a 2026 LiDAR scan, and vice versa, because both are evidence about the same persistent objects.

4.2 A worked cascade: when sound rewrites a photograph

The retroactivity is easiest to see in a case where the modalities are maximally unlike, because there the shared latent structure is the *only* thing that could possibly couple them. Consider an archive that has, over years, ingested thousands of photographs of a stone building from many angles. Pixel-level redundancy carries it only so far; the decisive collapse comes when the MDL ledger discovers that the cheapest encoding of the whole photographic corpus is not a library of 2D patches but a single 3D geometric template of the facade, against which each photograph is stored as camera pose, lighting parameters, and a small residual. This is itself a cross-view cascade — the geometry, once admitted, retroactively cheapens every prior photograph — but it remains within one modality.

Now a visitor uploads a video of a musician playing in the plaza. The microphone captures the direct instrument sound and, layered over it, the room’s acoustic impulse response: the reverberation of the facade. A modality-isolated codec would store these echoes as expensive, structureless audio residual. The shared library does something else. The audio matcher, hunting for any template that prices the echoes below their raw cost, finds the visual system’s 3D geometry already sitting in the library and discovers that a coarse acoustic ray-trace through that geometry predicts most of the reverberation. The audio file is encoded *against a template built from photographs*. Sound borrows geometry.

Then the informative failure. Suppose a sharp, distinct echo arrives that the geometry does not predict — a persistent anomaly in the audio residual, structured and stubborn. The matcher’s cheapest available explanation is a deep recessed alcove in the facade at a specific location, which would produce exactly that late reflection. Admitting the alcove dissolves the audio anomaly and pays its rent, so the shared geometry template is revised. But the geometry is shared, and its revision triggers the background process across the entire archive — including an eight-year-old photograph taken from an awkward angle that had always carried an unexplained dark patch exactly where the alcove now sits. For eight years that patch was stored as raw pixel residual, because nothing in the library cast it. Re-evaluated against the revised geometry, the patch is now the alcove’s shadow: the pixels are discarded, replaced by a pointer to the geometry, and the old photograph shrinks. A recording of a guitar has retroactively compressed a JPEG taken before the recording existed, because the archive treats both as evidence of one persistent latent world and the flow of correction runs in whichever direction the residual is largest. This is the cross-modal case of the sheaf failure of Chapter 2: a local section — the geometry that satisfied the photographs — failed to extend over the audio, and forced a gluing that then propagated back through everything the geometry touches.

4.3 The crossover point

There is a measurable moment in the life of a well-observed entity — a building seen by thousands of photographs, security footage, drone passes, LiDAR, drawings, maintenance logs — at which the archive’s model of the entity contains more information than any single recording of it, and the cheapest encoding of a *new* observation is “render from model, store the diff.” Past that crossover the archive has quietly changed kind: it is no longer a collection of files with a clever index but a persistent generative model for which recordings are evidence. The compressed representation of the corpus *is* the world model plus residuals,

$$\text{Data} \rightarrow \mathcal{M} + \text{Residuals},$$

and storage cost asymptotically tracks the world’s novelty rate rather than the sensors’ data rate. The world model here is not an abstract weight matrix: it is a *Holonomic Space* in the sense of the monograph of that name — a structure reconstructed entirely from local, recursive persistence constraints, with no global blueprint anywhere in the system, every part of which is recoverable from admissible paths through the rest.

The crossover deserves to be a theorem rather than an image. Let \mathcal{M}_n be the archive’s latent model of an entity after n observations O_1, \dots, O_n drawn through measurement kernels from a latent structure of finite description, and let $I(\cdot; \cdot)$ denote mutual information.

Theorem 4.1 (World-model crossover, schematic). *Suppose the observations are conditionally independent given the latent structure, each carries information about it bounded away from zero ($I(\mathcal{M}; O_i) \geq \epsilon > 0$ on a shared component), and no single kernel exposes the full structure ($I(\mathcal{M}; O_i) \leq C < H(\mathcal{M})$ for all i). Then there exists finite n^* such that for all $n \geq n^*$,*

$$I(\mathcal{M}; \{O_1, \dots, O_n\}) > \max_i I(\mathcal{M}; O_i),$$

and under diverse kernels $I(\mathcal{M}; \{O_i\}) \rightarrow H(\mathcal{M})$ while $\max_i I(\mathcal{M}; O_i)$ remains bounded by C . Past n^* the latent model is strictly the dominant information-bearing object in the archive, and the per-observation marginal storage cost converges to the conditional entropy $H(O | \mathcal{M})$ — the novelty rate.

The content is in the conditions: *diversity* of kernels (new angles, new modalities) is what drives the joint information past any single observation’s ceiling, which is why a LiDAR sweep retroactively cheapens photographs and why archives that ingest only one viewpoint never cross over. The crossover point n^* is measurable in a running archive — it is the moment “render from model, store the diff” wins the MDL comparison for typical new observations — so the metaphor of Chapter 1 has become an instrument reading.

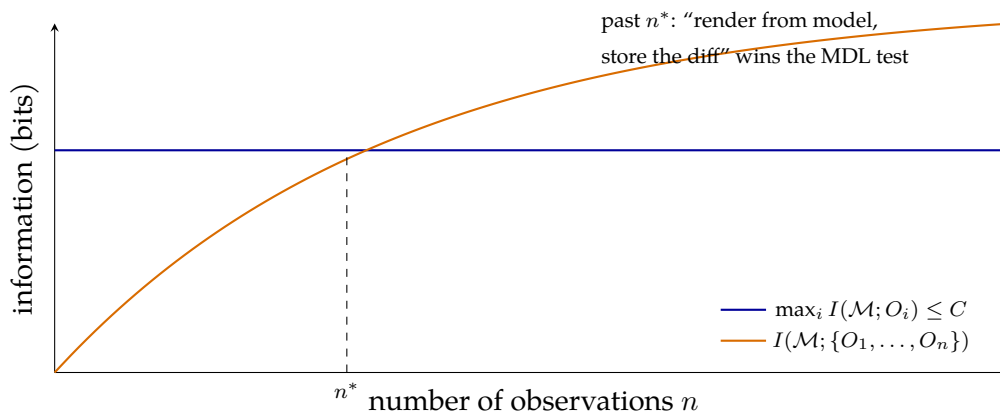


Figure 4.1: The world-model crossover. Any single observation’s information about the latent structure is capped by its measurement kernel (C), while the joint information of diverse observations climbs toward $H(\mathcal{M})$. Past n^* the latent model is the dominant information-bearing object and per-observation storage cost converges to the novelty rate $H(O | \mathcal{M})$.

Storage cost tracking novelty rather than data rate is also the engineering restatement of the predictive-processing view of memory: brains do not store retinal images; they store a latent room and update it, which is why familiar environments are cheap to perceive and anomalies are expensive, attention-grabbing, and memorable. Chapter 8 returns to this with the *Beyond Prediction Error* apparatus.

The industrial consequence of the crossover deserves to be stated plainly, because it overturns an assumption built into every storage system in use: that the cost of recording scales with the resolution of the sensor. Past n^* it does not. The storage cost of a stream decouples from the sensor’s data rate and tracks instead the *novelty rate* of the world the sensor observes — the conditional entropy $H(O | \mathcal{M})$ of the observation given the model.

A 16K camera left running in an empty, unchanging room generates an astronomical raw data rate and a novelty rate near zero: the archive pays once for the room's geometry and lighting on the first day, and thereafter stores almost nothing, because each frame matches the prediction and leaves no residual. It begins to spend bits again only when something unmodeled crosses the frame — a person, a moved chair — and even then only the bits that the existing model cannot render. Doubling the sensor resolution, in this regime, changes the storage bill almost not at all; what changes it is the arrival of the genuinely new. The byte is no longer the unit of measurement but the unit of surprise.

4.4 Residuals as the archive's frontier

In a mature multimodal archive the residual stream is the most interesting object in the system. Residuals are, by construction, whatever the world model cannot yet explain: sensor noise, genuine novelty, and — the critical third category — *systematic* misfit, the signature of a missing template. Monitoring residual statistics per region of template space gives the archive an intrinsic curiosity signal: persistent structured residual marks the place where the next collapse event is available. The archive's frontier of maximal compressible-but-uncompressed structure is its research program.

Part III

Information-Theoretic Foundations

Chapter 5

Two-Part Codes, MDL, and the Rent a Template Pays

5.1 Kolmogorov ideal and computable practice

The ideal compressed form of a corpus D is its Kolmogorov complexity $K(D)$: the length of the shortest program producing it. K is uncomputable and, worse, *unreachable* — no effective procedure converges to it with a certificate. The rematching archive's $\Lambda^*(\sigma)$ is the computable, situated surrogate: the best description reachable from the present library by admissible moves. The theory of the archive is therefore not approximation theory against K but *reachability theory* over description space: which representations are accessible from which, at what cost, through which intermediate states. Three consequences follow.

1. **Path dependence.** Two archives fed the same corpus in different orders inhabit different regions of description space and may converge to different local optima. The history of an archive is not noise around its content; it is constitutive of its representational capacities — event-sourced identity at the level of the whole system.
2. **No global optimality claims.** Every compression number the archive reports is indexed to its stratum: “4.1×, from here.”
3. **Discovery as path construction.** Improving compression is not search through a fixed space but *construction of new paths* — each admitted template is an edge that did not previously exist in the reachability graph.

5.2 MDL admission as Bayesian model comparison

The admission rule — admit t iff total description length drops — is the MDL form of a Bayes factor test, with $2^{-L(\mathcal{T})}$ the prior over libraries and $2^{-L(D|\mathcal{T})}$ the likelihood. The archive is performing continual Bayesian model refinement with the model class itself under construction. Two refinements matter in practice.

Deferred settlement. A template's value cannot be evaluated at admission time because its payoff arrives through future matches and cascades. The archive therefore posts templates on

credit: admitted provisionally, charged interest (their storage), and retired if, after a probation window, their realized savings have not covered cost. The library is a market of hypotheses, and RETIRE is bankruptcy.

Code-length smoothing. Reference costs are entropy-coded against template usage frequencies, which the cascades themselves change. Admission decisions therefore use shadow prices — expected post-cascade code lengths — rather than current ones, or the system oscillates.

5.3 Compression = prediction = comprehension

The arithmetic-coding identity $L(D | \mathcal{M}) = -\log_2 P_{\mathcal{M}}(D)$ makes compression and prediction the same quantity viewed from two ends: a model compresses exactly as well as it predicts. The rematching archive adds the third vertex of the triangle. Because its model class grows by *naming* recurrent structure (templates are discrete, referenceable, compositional entities), improvements in prediction are realized as improvements in an explicit ontology. Compression progress is therefore a metric of comprehension in a strong sense: the archive gets smaller precisely insofar as it has identified the entities and composition laws actually generating the data. The compression ratio is an instrument reading on the truth of the archive's ontology — noisy, local, and stratum-relative, but an instrument reading nonetheless.

5.4 The Weierstrass lesson: roughness, typicality, and forced strata

There is a nineteenth-century theorem that reads, in retrospect, like a warning addressed to every system built on the assumption that signals become simple when examined closely enough. Before the 1870s the working intuition of analysis held that continuity and smoothness were nearly the same property: a continuous curve was expected to be differentiable except at isolated singularities, because every curve one could draw looked locally straight under sufficient magnification. Weierstrass's function

$$f(x) = \sum_{n=0}^{\infty} a^n \cos(b^n \pi x), \quad 0 < a < 1, \quad ab \geq 1,$$

destroyed the intuition with two ingredients in deliberate tension: the factor a^n shrinks amplitudes fast enough to force uniform convergence and hence continuity, while b^n grows frequencies fast enough that new oscillations arrive at every scale. There is no magnification at which the graph becomes approximately straight; the derivative exists nowhere; the graph's dimension, $D = 2 + \log_b a$, sits strictly between curve and surface.

The function matters to this monograph for three distinct reasons, in ascending order of consequence.

First, it marks a singular point in description space. The Weierstrass function is maximally rough as a trajectory and nearly minimal as a program: its Kolmogorov complexity is that of a one-line formula, while its ontological deficit $\delta_{\mathcal{T}}$ under any finite-depth fragment library is severe. An archive ingesting samples of f at increasing resolution would find that the template hierarchy never bottoms out — every level of fragments exposes fresh oscillatory residual beneath it, since $ab \geq 1$ guarantees structured mass at all scales. Yet the signal is one of the cheapest objects in existence *provided the right template kind is admissible*: self-similarity

means the same generator recurs at every level with rescaled parameters, which is precisely the regularity fractal compression exploits. No accumulation of template *instances* reaches the short description; only the admission of a recursive, self-referential template *kind* does. The Weierstrass function is thus a forced stratum transition expressed as a single formula — the cleanest possible example of a representation that exists, is short, and is unreachable from an entire class of libraries.

Second, it exhibits the failure mode of local frames. In the admissibility reading, differentiability is local predictability: a tangent direction is an admissible local flow, a statement that the trajectory's immediate future is reachable by linear extrapolation from its present. The Weierstrass function is continuous in position but discontinuous in direction: the graph stays connected while the assignment of tangent directions exists nowhere. The configuration space is intact; it is the fibration over it — the frame bundle every locally linear method silently requires — that is empty. Level-1 matchers with shift-gain-warp invariance groups are locally linear methods in exactly this sense, and the function is the proof that their applicability is a property of the source, not of signals as such. In RSVP terms the rematching flow \mathbf{v} cannot stabilize on such a signal at any finite scale: every relaxation step receives fresh Φ arriving from below, and only a scale-coupling term in the constraint geometry — a recursive template — closes the cascade.

Third, and most seriously, the monster is generic. The lasting shock was not Weierstrass's example but the later discovery of its typicality, and the typicality holds in every inequivalent sense of "most" that has been tried: nowhere-differentiable functions are residual in $C[0, 1]$ (Banach and Mazurkiewicz, 1931), prevalent in the measure-theoretic sense appropriate to spaces carrying no translation-invariant measure, and probabilistically the default — Brownian sample paths are almost surely nowhere differentiable, so nature's canonical random curve is a Weierstrass-type object. Smooth curves are a measure-zero island. The correct conclusion is not that the world is hostile but that human geometric intuition was trained on that island: the curves an arm can draw are the curves a physical process with bounded jerk can produce. Smoothness was never a property of curves; it was a property of *our generators*.

For the archive this is the decisive point, because it dissolves an objection and installs a discipline. The objection: if generic signals are incompressibly rough, the whole program is hopeless. The dissolution: the archive never claimed to compress generic signals — no system can, and Shannon's counting argument already says so for generic bitstrings. The archive compresses the outputs of *the world's generators*, and the Weierstrass lesson, run in reverse, says that every regularity the library accumulates is a discovered fact about those generators, never a fact about signal space. The template hierarchy is an inventory of the measure-zero island we actually inhabit. The discipline follows: roughness in the residual stream is informative in a way smoothness never is. Residual that stays structured across scales — the $ab \geq 1$ signature — is the archive's evidence that it faces a generator with scale coupling, and that the profitable admission is a recursive kind rather than more instances. Turbulence, $1/f$ noise, financial series, fracture surfaces, and coastlines all present it. A mature archive should treat scale-stable residual texture as the single strongest anomaly class in its TARTAN tiling: the place where the next stratum is not merely available but *named in advance* by a theorem from 1872.

Chapter 6

The Thermodynamics of Rematching

6.1 Description length as free energy

Write the archive’s objective as

$$\Lambda = \underbrace{L(\mathcal{T})}_{\text{model cost}} + \underbrace{\sum_o L(o | \mathcal{T})}_{\text{misfit}},$$

formally a free energy: model cost plays entropy’s role (capacity spent on structure), misfit plays energy’s (unexplained constraint violation). Relaxation dynamics that lower Λ through local moves, with occasional uphill admissions crossing barriers into deeper basins, give the archive the phenomenology of an annealing glass: fast quench on ingestion (greedy encode), slow structural relaxation in the background (rematching), aging (old regions improve as the library matures), and avalanches (collapse events) with heavy-tailed sizes. The glass picture is not decoration; it predicts engineering facts — that rematching yield per cycle decays as a power law between discoveries, that “reheating” (loosening match thresholds temporarily) escapes local minima, and that archives quenched too aggressively at ingestion develop frozen-in defects (bad early segmentations) that later cascades cannot cheaply repair.

6.2 RSVP: the field-theoretic reading

The RSVP framework describes systems by a scalar density Φ , a vector flow \mathbf{v} , and an entropy field S relaxing under coupled dynamics. The rematching archive instantiates the triple exactly, and the identification is more than notation:

- Φ is *unexplained structure density* over description space — where residual mass sits, weighted by its systematicity. High Φ marks compressible-but-uncompressed regions: the anomaly field.
- \mathbf{v} is the *rematching flow* — the directed transport of description mass through the reachability graph as rewrites propagate. Collapse events are shock fronts in \mathbf{v} : discontinuous bulk transport triggered when local flow exceeds a threshold.
- S is the *realized code* — entropy already extracted into efficient form; the archive’s crystallized understanding.

The triple, however, is only the visible portion of the plenum architecture, which is a five-engine system. The archive supplies the remaining two engines as naturally as the first three. The fourth is the *admissibility constraint surface* itself — the geometric engine: the metric and barrier structure on description space, determined by the current template DAG, that prices every move and thereby regulates the rate and direction of relaxation. It is not a backdrop to the flow; it is dynamical, since every ADMIT and REFACTOR deforms it, and most of the system’s interesting behavior (plateaus, barriers, lock-in) is curvature of this surface. The fifth is the *novelty source* — the injection engine: ingestion deposits fresh Φ at the boundary of the complex at the world’s novelty rate, and RETIRE provides the conjugate sink. With all five engines named — density, flow, crystallized entropy, constraint geometry, source/sink — the archive completes the architectural pentad rather than sampling it.

The translation can be made exact by writing the coupled relaxation equations. Let $\Phi(x, t)$ be unexplained structure density at coordinate x in the description complex and $\mathbf{v}(x, t)$ the rematching velocity field transporting it toward realization. Conservation of description mass with conversion into crystallized code reads

$$\frac{\partial \Phi}{\partial t} + \nabla \cdot (\Phi \mathbf{v}) = -\Gamma(\Phi, \mathcal{T}) + \eta(x, t), \quad \frac{\partial S}{\partial t} = \Gamma(\Phi, \mathcal{T}),$$

where η is the novelty source (ingestion) and $\Gamma(\Phi, \mathcal{T}) \geq 0$ is the *rate operator*: the speed at which anomalous density is crushed into structural entropy, a function of both the local residual mass and the library’s current capacity to express it. Γ vanishes on a plateau not because Φ is zero but because the constraint surface offers no admissible channel — the stratum is in quasi-equilibrium with positive trapped density, which is exactly the ontological deficit $\delta_{\mathcal{T}}$ in field form. A collapse event is a shock front: an ADMIT removes a structural bottleneck in the template DAG and

$$\nabla \cdot \mathbf{v} \longrightarrow -\infty \quad \text{on the front,}$$

a catastrophic local convergence of flow as trapped mass from a wide basin funnels through the newly opened channel into S .

Relaxation thus converts Φ into S along \mathbf{v} , subject to the admissibility constraints of the current stratum. Plateaus are quasi-equilibria where \mathbf{v} has decayed because no admissible move moves mass; an admission perturbs the constraint surface, reopens channels, and the field avalanches into the next basin. The wave structure of compression is, in RSVP terms, the generic behavior of entropy-relaxation in a stratified constraint field — the same formal behavior the framework assigns to cosmological smoothing, neural settling, and fiscal crisis. The archive is a tabletop RSVP system with every field directly measurable, which makes it something the broader program has lacked: an *experimental apparatus*.

6.3 TARTAN and the texture of the residual field

TARTAN’s recursive tiling supplies the spatial discipline: description space is tiled at multiple scales, each tile carrying local residual statistics, and the rematching scheduler is a TARTAN traversal — attention flows to tiles whose residual texture is most structured. Curiosity, in this system, is literally a tiling-guided search policy over the Φ field.

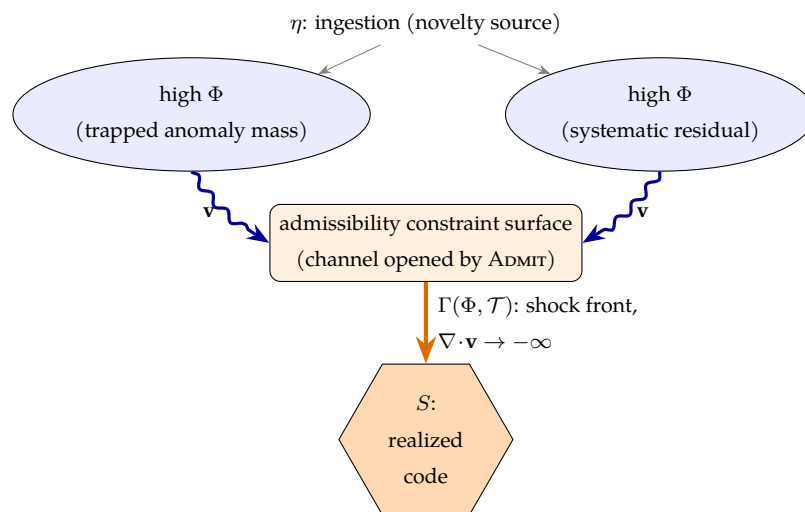


Figure 6.1: The archive as a tabletop RSVP system. Ingestion (η) deposits unexplained structure density Φ ; the rematching flow \mathbf{v} transports it through channels priced by the admissibility constraint surface; the rate operator Γ crushes it into crystallized code S . On a plateau Γ vanishes with $\Phi > 0$ — the ontological deficit in field form. An ADMIT removes a bottleneck and the trapped mass funnels through as a shock front: a collapse event.

Part IV

Admissibility, Strata, and Frozen Processes

Chapter 7

The Archive as a Stratified Reachability Space

7.1 Strata of description

Define an equivalence on archive states: $\sigma \sim \sigma'$ iff each is reachable from the other by moves of bounded (cheap) cost. The equivalence classes are *strata*; within a stratum the archive rearranges freely, between strata it must pay — an admission that temporarily inflates Λ , a refactoring campaign, a re-segmentation of frozen early defects. The compression staircase is the trace of stratum transitions: plateaus are life within a stratum; jumps are crossings.

This is the frozen-process ontology made operational. What the archive *is*, at any moment, is not its bitstring but its position in the stratified reachability space: which descriptions it can reach cheaply (its fluid degrees of freedom), which it can reach only by crossing barriers (its frozen ones), and which are not reachable at all from here (not part of its present ontology in any operative sense). A template hierarchy is a frozen process: the residue of past relaxation that now constrains, enables, and prices all future relaxation. The primitive fact about a representation is never that it exists in some platonic codebook; it is whether and at what cost it is reachable.

7.2 Persistent anomalies and forced revision

A *persistent anomaly* is residual mass that no within-stratum move reduces: structure visibly present (systematic, recurrent, sketch- correlated) yet inexpressible over the current library at profit. The geometry of *Persistent Anomalies and the Geometry of Ontology Revision* applies verbatim, and the sheaf reading of Chapter 2 makes it topological: a structured residual is an obstruction that cannot be smooth-mapped onto the current template library without tearing the existing reference DAG. The anomaly defines a misfit gradient pointing out of the stratum; the set of admissions that would dissolve it defines the local geometry of ontology revision; and the archive's policy question — when to stop patching residuals and admit a new template kind — is exactly the philosopher-of-science's question of when accumulating anomalies force a paradigm change. The archive answers it with an MDL ledger rather than a crisis of confidence, which suggests, in the other direction, a ledger-theoretic reading of

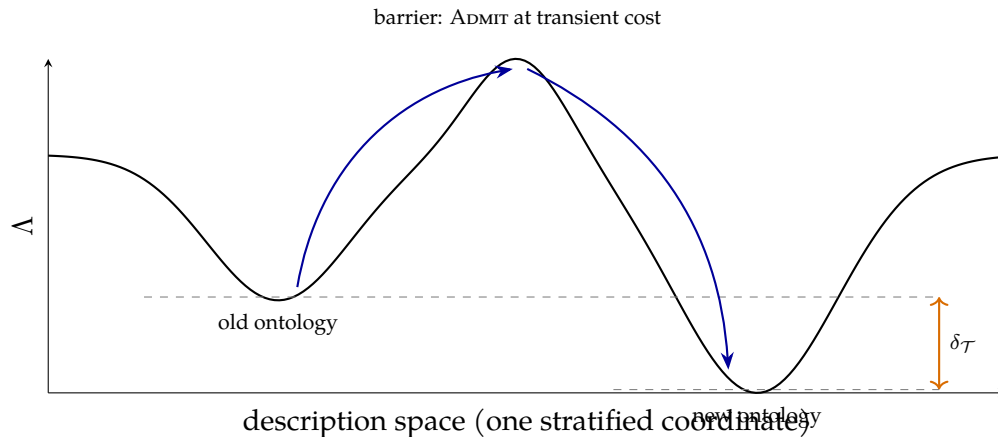


Figure 7.1: Strata and the ontological deficit. Within-stratum moves relax Λ to the local floor; the deficit $\delta\tau$ is the gap between that floor and the one reachable only by crossing the barrier — an admission whose cost is paid up front and repaid by the cascade. Learning is not smooth search; it is the construction of admissible paths across the barrier.

theory change itself.

The qualitative analogy supports a quantitative claim.

Theorem 7.1 (Ontology revision as phase transition, schematic). *Let trapped anomaly mass $\delta\tau$ within a stratum grow with ingestion while no admissible within-stratum move reduces it. Define the order parameter as realized-code fraction $S/(S + \Phi)$. Then the admission of a template family that dissolves the anomaly is a first-order transition: the order parameter jumps discontinuously by the collapse magnitude m , latent “heat” m is released as code-length savings in a cascade of finite duration, the two strata coexist during the cascade (rewritten and not-yet-rewritten regions), and hysteresis obtains — the reverse transition (retiring the family) does not occur at the same ledger threshold, because the rewritten archive now depends on it.*

In the archive this is not philosophy. It is an observable event with a timestamp in the Admissibility Log, a measurable $\Delta\Lambda$, and a front velocity through the reference DAG. Scientific revolutions have long been said to resemble phase transitions; here is a system in which the resemblance is an identity, and in which Kuhn loss is measurable too: structure expressible in the old stratum that the new one prices higher, visible as residual mass that *increases* locally across the transition.

7.3 Belief geometry of the encoder

The matcher’s internal state — which template a segment is about to be read as — moves through a space with the structure described in *Belief Geometry and Reachability*: basins around templates, ridgelines where readings are contested, and admissibility constraints determining which reinterpretations are reachable from a given partial parse. Rematching is belief revision performed on an archive’s scale: the cascade after an admission is the propagation of a changed reading through everything that reading touches. The transformer-residual-stream

analogy is direct, and the archive again plays apparatus: parse basins and their boundaries can be enumerated, measured, perturbed.

Chapter 8

Memory, Brains, and Resonance Without a Substrate

The claim that this architecture mirrors biological memory can be made with some precision. Predictive-processing accounts hold that cortex maintains generative models and propagates only prediction errors; *Beyond Prediction Error* extends this to constraint-relaxation rather than error-minimization proper, and the archive sides with the extension: its dynamics minimize a two-part description length, not a pointwise error, and its “errors” (residuals) are first-class stored objects with their own economy, not transient signals to be silenced. Memory consolidation maps onto background rematching — experiences ingested verbatim and gradually rewritten as references to schemas, with sleep as the maintenance window for cascade processing. Reconsolidation maps onto rewrite-on-recall. Semantic memory is the template hierarchy; episodic memory is the residual stream plus the event ledger; the reminiscence bump and schema-dependent forgetting fall out of rent-and-retirement economics. *Resonance Without a Substrate’s* thesis — that memory is maintained pattern-reachability rather than stored substance — is here an implementation note: nothing in the archive is stored “as itself”; everything persists as the capacity to be regenerated along an admissible path, and what is preserved across rewrites is reachability of the observation, not any particular bit pattern. Identity without substrate, kept honest by the Spherepop ledger.

Two consequences of this mapping are worth drawing out, because they are where the architecture stops being an analogy for memory and starts predicting its known peculiarities. The first is that recall is rendering, not playback. If episodic memory is a sparse ledger entry — a timestamp, a set of semantic templates to load, and a thin residual of what those templates failed to predict — then remembering is the act of re-running those templates to regenerate the scene, and the templates being run are the *current* ones, not the ones in force when the memory was laid down. A memory recalled in 2026 is rendered through 2026’s hierarchy. This is, in the archive, the exact mechanism of false memory: not corruption of a stored recording but faithful rendering from a library that has itself been rewritten by every collapse event since the encoding. The unreliability of human memory is not a flaw bolted onto an otherwise veridical store; it is the unavoidable signature of a store that holds reachable paths rather than substance. A system that compressed by understanding could not behave otherwise.

The second is the phenomenology of recognition, which the predictive account renders almost tactile. Meeting a friend of twenty years, one does not process the geometry of their

face; recognition loads a deep template and what reaches awareness is the residual — they look tired, they have new glasses, they have aged since last time. Perception spends its bits on the delta and nothing on the prediction. This is the empty-room argument of Chapter 4 run inside a skull: a familiar environment is cheap to perceive precisely because it leaves almost no residual, and the cost of an experience is its novelty rate, not its sensory bandwidth. The move of sudden understanding — the click when a hard concept resolves and abruptly feels weightless — is, in the same currency, a collapse event in one's own template hierarchy: a mass of trapped residual finding a single template that explains it, the description length dropping, the cognitive load lifting. The phenomenology of comprehension is the phenomenology of compression because, in this monograph's central claim, they are one process.

The closing thought is the one that follows from taking the mapping seriously rather than decoratively. If memory is a rematching archive, then a childhood memory has not lain frozen for decades, degrading like a photograph in a drawer. It has been continually re-rendered: every year of additional understanding — of people, of loss, of physics, of how the world works — has given the background process new templates against which to re-evaluate that afternoon's thin residual, and the memory has been quietly rewritten each time into the vocabulary of a later self. One does not remember the event. One renders the most recent, most heavily refactored template of it. The past, in a system that compresses by understanding, is not fixed; it is the thing that keeps collapsing into a sharper and more efficient story, and the collapses do not stop while the archive is still learning.

Chapter 9

Threads, Projections, and the Category of Admissible Paths

9.1 Against modality-first typing

How should cross-modal templates be typed? The conventional answer is algebraic and modality-first: distinct categories `Image`, `LiDAR`, `Audio`, with hand-specified bridges,

$$\text{Image} \rightarrow \text{Geometry}, \quad \text{LiDAR} \rightarrow \text{Geometry}, \quad \text{Audio} \rightarrow \text{Material}.$$

This is how multimodal systems are built today, and the entire arc of this monograph argues against its ontology. The archive does not fundamentally contain images, audio streams, or point clouds; it contains reachable descriptions, and the modalities are measurement kernels. The primitive object of the type system should therefore not be a modality but a *process thread*.

9.2 Yarncrawler threads as the primitive type

A Yarncrawler thread is neither an image nor a scan. It is a trajectory through latent process space that multiple sensors can partially observe. Typing a template means assigning it

$$\tau : \mathcal{P} \longrightarrow \mathcal{O},$$

a map from a process manifold to an observation manifold, and a modality is a projection operator applied to the thread:

$$\pi_{\text{img}}(\tau), \quad \pi_{\text{lidar}}(\tau), \quad \pi_{\text{audio}}(\tau).$$

One thread satisfies many observation types simultaneously, and the conventional type-mismatch question dissolves. One never asks whether an image template can substitute for a LiDAR template; one asks whether two observations are projections of the same thread — a *reachability* question, settled by the ledger, not a typing question settled by fiat.

This is where CLIO enters structurally rather than decoratively. The CLIO projection $\pi : X \rightarrow M$ already formalizes the distinction between an underlying structure and its observed manifestation. A cross-modal template is a persistent element $\tau \in M$ of the latent manifold; modalities are projection families; and the type discipline reduces to admissibility constraints governing which projection families may be attached to which threads.

9.3 Geometric and algebraic at once

Must the process manifold be geometric or algebraic? It must be both, and the model is the path class of algebraic topology, which is simultaneously a geometric trajectory, an algebraic object, and a compositional entity. A thread's geometric structure determines which transformations of it are reachable — distances, barriers, collapse basins, the whole RSVP machinery; its algebraic structure determines how it composes — so that

$$\text{Building} = \text{Foundation} \circ \text{Walls} \circ \text{Roof}$$

is well-formed while every factor retains a position in reachability geometry. Without the geometry there is no admissibility; without the algebra there is no template construction. The natural formal home is a category whose objects are strata of a process manifold and whose morphisms are admissible path classes, with observation modalities as functors to the respective measurement categories and collapse events as topology-changing reductions in path complexity. Open Problem 3 is restated in these terms among the open problems of Chapter 12, and the restatement has a deflationary punchline: formulated this way, the compression algorithm nearly disappears. What remains is a theory of how observations become organized into persistent process threads — with compression surviving as the numerical signature that the organization is improving.

9.4 The loop foundation models do not close

Contemporary foundation models exhibit a weak form of the archive's dynamic: as training proceeds, structure initially carried by many parameters becomes expressible through increasingly reusable latent features. But training is one-way — the model's representation changes while the corpus is never rewritten. The rematching archive closes the loop: the model improves, the archive rewrites itself in the improved vocabulary, and the rewritten archive changes which further models are reachable. Memory and representation become a coupled dynamical system rather than a frozen corpus and a plastic model, and the belief-geometry results of the previous chapter acquire a second life: the residual-stream basins of a trained model are a snapshot of exactly the structure the archive holds as a live, revisable ledger.

Chapter 10

Compression as Epistemology

10.1 The archive as a scientific community

Assemble the pieces: hypotheses posted on credit, settled against an evidence ledger; an ontology of named entities under perpetual revision; normal-science plateaus and revolutionary collapses; anomalies as the engine of change; path-dependent, stratum-relative knowledge with no view from nowhere; and a single scalar — total description length — playing the role of empirical adequacy. The rematching archive is a small philosophy of science that runs. Its lessons cut both ways. Toward engineering: theory-change phenomena (Kuhn loss, incommensurable re-segmentations, degenerating research programs as template lineages that no longer pay rent) are predictable failure and maintenance modes of large archives. Toward philosophy: MDL-with-deferred-settlement is a candidate mechanism for rational paradigm change that needs no appeal to taste or mob psychology — revisions happen when the reachable two-part code on the other side of the barrier is shorter, and the sociology is in the barrier-crossing schedule, not the criterion.

10.2 Knowledge as navigation

The deepest integration with the surrounding program is also the simplest to state. Knowledge, throughout these frameworks, is navigation through reachable states; understanding is the growth of reachability. The archive renders this literal: a signal is known to the degree that it is reachable from a short description over the current hierarchy; learning is the construction of new admissible paths; and the punctuated collapse of the archive's size is the visible signature of new regions of description space becoming navigable. Compression is not a thing one does to knowledge. Compression is what knowledge looks like from the storage layer.

Part V

Consequences, Designs, and Open Problems

Chapter 11

Engineering Consequences

11.1 What changes when size is not a property of files

In a rematching archive the compressed size of an object is a function of the whole system's history and present library. Consequences: storage accounting must be marginal, not per-file (an object's cost is $\Lambda(\text{archive}) - \Lambda(\text{archive} \setminus o)$, which can be near zero for well-modeled content and can *decrease over time*); deduplication is subsumed (duplicates are the degenerate case of templates); deletion is nontrivial (removing evidence can orphan templates and *inflate* the remainder — the ledger must price this); and provenance is mandatory, since bits are rewritten but identity must persist. Decoding remains cheap and local — follow references down the DAG — which is the saving grace: all the intelligence is on the write path.

The marginal definition has a consequence sharp enough to disorient. The size of a file is not a property the file carries; it is a property of the file's relationship to everything already known. The same photograph of a cat costs ten megabytes in an empty archive that must build the animal's geometry from nothing, and a few bytes in an archive holding a million prior images of that same cat, where it resolves to a pose, a lighting parameter, and a thin residual against an existing model. "How big is this file" is no longer a well-posed question without specifying the archive; the operating system's confident report of 2.4 megabytes is an artifact of isolation, true only because the file is stored alone. Once storage is relational, the entire vocabulary of conventional file management is unsettled: deduplication, the practice of scanning a server farm for byte-identical copies to delete, becomes vacuous because identity of content is the easiest case the template matcher already handles before any duplicate is written. The hard cases are not identical files but near-misses, and those the archive handles by gradient, not by equality test.

11.2 Failure modes the theory predicts

The frameworks predict the pathologies. *Template capture*: an early bad segmentation becomes load-bearing and resists revision — frozen defects, requiring scheduled reheating. *Hallucinated economy* in lossy regimes: an overconfident world model spends residual bits poorly and reconstructions drift plausible-but-wrong — the distortion lattice must retain enough residual to keep the model falsifiable, an empirical-adequacy constraint, not a fidelity knob. *Ontology lock-in across modalities*: a geometry template family that explains images well can suppress

admission of a better family serving LiDAR — stratum competition, requiring per-modality anomaly audits. Each failure mode is a known phenomenon of theory change wearing a systems costume.

Two of these have exact analogues outside computing that are worth naming, because they show the pathologies are not artifacts of the implementation but generic features of any system that compresses by accreting structure. Template capture is path dependence with a sunk cost, and its canonical human form is the persistence of an inferior standard: a layout devised to slow nineteenth-century typewriters survives on devices that cannot jam, not because it is good but because the cost of unwinding the muscle memory and infrastructure built atop it exceeds the benefit of any superior replacement. The archive falls into the same trap whenever a flawed early template becomes load-bearing — millions of encodings point to it, and the bill for re-deriving them all under a better template is never worth paying, so the system stays in a local minimum it can see past but cannot afford to leave. Ontology lock-in across modalities is, in turn, the sociology of scientific resistance running autonomously: an entrenched template family competes for the same finite computational budget as the superior newcomer and, by being already woven through the hierarchy, can starve the better family of the cycles it needs to establish itself — the old paradigm defending its territory not from conviction but from incumbency, exactly as a generation of practitioners whose careers rest on the old model can delay the new one without anyone deciding to. That a CPU-scheduling contest reproduces the texture of academic inertia is not a metaphor laid on top; it is the same dynamic, which is why the same remedy — forcing the system to keep auditing the anomalies the reigning ontology would prefer to ignore — works in both.

Pathology	Theory translation	Mitigation
Template capture	An early bad segmentation becomes frozen and load-bearing; later cascades cannot cheaply repair it.	Scheduled reheating: temporarily loosen match thresholds to escape the local minimum, then re-anneal.
Hallucinated economy	An overconfident lossy world model drifts into plausible-but-wrong reconstructions.	Maintain the distortion lattice; keep residual retention above the empirical-adequacy floor so the model stays falsifiable by its own archive.
Ontology lock-in	One modality's successful template family suppresses admission of a better family serving another modality.	Per-modality anomaly audits: residual statistics tracked and thresholded separately for each measurement kernel.

Table 11.1: The three predicted pathologies of a rematching archive, each a known phenomenon of theory change in a systems costume, with its engineered mitigation.

11.3 A minimal experimental program

The apparatus claim should be cashed out. A reference implementation — single-modality audio, level-1 fragment matcher with LSH sketches, Re-Pair-with-MDL grammar above it,

ledgered rewrites — suffices to measure: the collapse-size distribution (test the power-law conjecture); aging curves of rematching yield; staircase shape versus corpus order (path dependence); and anomaly-mass dynamics around forced admissions. Every quantity in the RSVP reading (Φ as residual density, \mathbf{v} as rewrite flux, S as realized code) is directly loggable. The author knows of no other setting where the full triple is observable at native resolution.

Chapter 12

Open Problems

1. **Convergence across orderings.** Characterize corpora for which all ingestion orders reach the same stratum, and those exhibiting permanent path-dependence. Conjecturally this is a property of the generative grammar’s ambiguity structure.
2. **Optimal speculation.** The admission-on-credit policy is a bandit problem over an evolving action set whose payoffs arrive through cascades. Regret bounds are open even in toy grammars.
3. **The category of admissible process paths.** Sharpening Chapter 9: can a multimodal archive be formalized as a category of admissible process paths on a stratified manifold, where observations are functorial projections of Yarncrawler threads, the CLIO map $\pi : X \rightarrow M$ supplies the latent/manifest distinction, and collapse events correspond to topology-changing reductions in path complexity? What is gained and lost relative to modality-first typing?
4. **Distortion and truth.** In the lossy multimodal regime, formalize the empirical-adequacy floor: the minimum residual retention under which the world model remains falsifiable by its own archive.
5. **Thermodynamic cost.** Landauer accounting of rematching: collapse events erase representational distinctions and must dissipate; estimate the joules per bit of understanding and compare against the glass picture’s predictions for avalanche statistics.
6. **The crossover theorem, rigorously.** Chapter 4 states the world-model crossover schematically; prove it with explicit rates — n^* as a function of kernel diversity, latent dimensionality, and noise — and characterize corpora that never cross over.
7. **Conservation of transformability.** Formalize the relationship between an archive’s compression depth and its remaining path-flexibility. Does a high ratio achieved through deep template nesting minimize the number of future admissible paths (ontological lock-in), or does the extraction of genuine regularities maximize the navigable future trajectories of the residual stream? Conjecturally both, on different timescales — which would make transformability a conserved-then-dissipated quantity of rewrite cascades. Map the question onto the manifold interpretation of ecological network resilience, where the same tension between efficiency and adaptability is well studied.

Coda

A codec is a promise to forget nothing while understanding nothing. An archive that rematches makes the opposite promise: it will understand more and more, and what it stores will weigh less and less, and the two facts will be the same fact. Its history is written in a staircase — long patient floors of normal encoding, then the sudden vertical of a collapse, when some quiet template turns out to be the name of half the past. Nothing was deleted in that moment. The world simply became reachable from a shorter sentence.

Bibliographic Notes

The classical background is Shannon’s source coding theorem and its realization in arithmetic coding; LZ77/LZ78 and their dictionary descendants; grammar-based compression in Sequitur (Nevill-Manning and Witten) and Re-Pair (Larsson and Moffat); vector quantization (Gersho and Gray); sparse coding (Olshausen and Field); and fractal block coding (Barnsley, Jacquin). Minimum description length follows Rissanen, with the two-part code and its Bayesian reading per Grünwald. Kolmogorov complexity and the uncomputability results are standard from Li and Vitányi. The compression-prediction identity underlies the Hutter Prize program and modern neural compressors. Predictive processing draws on Rao–Ballard, Friston, and Clark; consolidation and reconsolidation on the systems-memory literature. Kuhnian structure, Lakatosian research programs, and the anomaly-driven account of revision are engaged throughout Part IV. Framework references — RSVP, TARTAN, CLIO, Spherepop, Yarncrawler, the admissibility/reachability program, *Frozen Processes*, *Persistent Anomalies and the Geometry of Ontology Revision*, *Belief Geometry and Reachability*, *Beyond Prediction Error*, and *Resonance Without a Substrate* — are to the author’s monograph series (Flyxion, 2024–2026).