Ordered Patch Theory

Appendix T-9: Maintenance Cycle, MDL Pruning, and Recovery Conditions

Anders Jarevåg

May 11, 2026 | DOI: 10.5281/zenodo.19300777

Original Task T-9: Maintenance Cycle and Recovery Apparatus Problem: The main paper §3.6.3–§3.6.6 defines equations T9-1 through T9-13 (the Maintenance Cycle operator $\mathcal{M}_\tau$, MDL pruning Δ_{MDL}, consolidation gain $\Delta K_{\text{compress}}$, REM importance weighting $w(b)$). Appendices T-12 (Narrative Drift) and T-13 (Action-Drift) cite this apparatus as load-bearing. The framework lacks a consolidating appendix that (i) names the formal primitives explicitly, (ii) distinguishes the four pruning modalities that the main paper's $\Delta_{\text{MDL}} < 0$ leaves implicit, (iii) defines the recovery condition, and (iv) provides a stable formal target for the corollary appendices to reference. T-9 fills that gap. **Deliverable:** Consolidating appendix at the same epistemic tier as T-2 / T-15 (structural correspondence, not closed theorem). New content beyond the main paper: explicit predictive-gain definition $G_i(t, \tau)$, maintenance-cost decomposition with resource-capacity primary, four-pruning-modality distinction, recovery condition, corollary chain.

Closure status: STRUCTURAL CORRESPONDENCE (same tier as T-2 / T-15). This appendix is *not* a closed theorem appendix. It consolidates the Maintenance Cycle apparatus already operating in preprint §3.6 and adds four pieces of formal content the main paper does not carry: explicit predictive gain, resource-capacity cost framing, four pruning modalities, and recovery conditions. The §2 OpenAI-review caveats are observed: (i) the pruning threshold is presented in the form that coordinates with T-12's pending channel-independence reformulation (Phase 4); (ii) the existing main-paper equations T9-3 / T9-4 are *preserved as cited*, with T-9 introducing the resource-capacity refinement as an additional formal layer rather than silently changing the cited forms; (iii) resource-capacity cost is primary, with K-complexity as a structural-correspondence approximation. Open edges (§9): the resource-capacity vs. K-complexity bookkeeping needs full reconciliation with T-12 once T-12's reformulation lands.

§1. Setup — Active Model Components

The codec K_θ comprises a collection of *active model components* $\{\theta_i\}_{i \in I}$, where each θ_i is an addressable structural unit of the codec — a generative prior, a learned feature detector, a recurrent stack, a long-range coupling, or any other primitive that participates in producing the codec's predictions π_t and update operator $\mathcal{U}$ over time. The collection $\{\theta_i\}$ is finite at any given moment but can be expanded through consolidation (Pass II, preprint §3.6.4) or contracted through pruning (Pass I, preprint §3.6.3).

For the purposes of T-9, the components are taken as defined: T-9 does not derive what makes one θ_i vs another a "natural" component, which is a representational-learning question outside OPT's scope. The Maintenance Cycle apparatus operates on whatever decomposition the codec admits.

The Maintenance Cycle operator $\mathcal{M}_\tau$ (preprint Eq. T9-2) acts on the Phenomenal State Tensor $P_\theta(t)$ during low-load intervals ($R_{\text{req}}(t) \ll C_{\max}$). T-9 unpacks the three passes (pruning, consolidation, forward-fan sampling) into the explicit formal primitives §2–§6 below; the corollary chain in §7 then traces Narrative Drift (T-12) and Action-Drift (T-13) through these primitives.

§2. Predictive Gain $G_i(t, \tau)$

The predictive gain of a component θ_i over a window of length τ measures how much that component contributes to the codec's predictive performance on the input stream, holding the other components fixed:

$$G_i(t, \tau) \;:=\; I(\theta_i \,;\, X_{t+1:t+\tau} \mid \theta_{-i}) \qquad \text{(T9.2-1)}$$

where θ_{-i} denotes the rest of the codec without θ_i, and $I(\cdot;\cdot \mid \cdot)$ is conditional mutual information. The conditional form is essential: it isolates θ_i's *marginal* predictive contribution rather than its joint contribution with overlapping components.

Comparison to main paper Eq. T9-3. The main paper's MDL pruning quantity is

$$\Delta_{\text{MDL}}(\theta_i) \;=\; I(\theta_i \,;\, X_{t+1:t+\tau} \mid \theta_{-i}) - \lambda K(\theta_i) \qquad \text{(T9-3, preprint §3.6.3)}$$

T-9 names the first term $G_i(t, \tau)$ explicitly so the predictive-gain primitive can be referenced separately from the threshold-form pruning condition. This is purely a notational consolidation; the inequality preserved.

Window length τ. The predictive gain depends on the window length. Short τ captures fine-timescale prediction (motor control, working memory); long τ captures structural prediction (semantic regularities, narrative coherence).

The Maintenance Cycle Pass I pruning is evaluated at the longer-τ regime where genuinely-useless components have $G_i \to 0$. Pass II consolidation, by contrast, optimises over the short-τ regime where redundancy across overlapping components becomes salient.

§3. Maintenance Cost C_i — Resource-Capacity Primary

The maintenance cost of a component θ_i has two compatible formulations.

Form 3.1 — Resource-capacity (primary for T-9). The component's cost is the resource capacity it occupies in the codec's operational substrate:

$$C_i \;:=\; c_i^{\text{params}} + c_i^{\text{memory}} + c_i^{\text{compute}} + c_i^{\text{channel}} \qquad \text{(T9.3-1)}$$

where the four budgets are: parameter slots (number of weights or connections); memory footprint (in bits stored); compute cost (in operations per cycle); and channel capacity (bits of bandwidth the component consumes at the Markov-blanket boundary $\partial_R A$). Each c_i is observable in principle — for biological codecs via metabolic and physiological measurement, for synthetic codecs via direct instrumentation.

Form 3.2 — K-complexity approximation. The main paper Eq. T9-3 uses $\lambda K(\theta_i)$ where $K(\theta_i)$ is the prefix Kolmogorov complexity of the component:

$$C_i^{\text{K-approx}} \;:=\; \lambda \cdot K(\theta_i) \qquad \text{(T9.3-2)}$$

This is a *structural-correspondence approximation*: K-complexity is upper-semicomputable and not strictly additive across components (deleting one component may not reduce shortest-description length by its standalone $K(\theta_i)$, since components can share structure). The resource-capacity form (T9.3-1) is therefore primary for operational claims; the K-complexity form is retained for theoretical analyses where the additivity approximation is acceptable.

Why two forms. The OpenAI review of T-12 (appendix-corrections memo §2.8) correctly noted that K-complexity is not additive across components and recommended resource-capacity measures for operational claims. T-9 adopts resource-capacity as primary but preserves the K-complexity form because the existing main-paper Eq. T9-3 and the T-12 Theorem T-12 proof both cite the K-complexity form. The resource-capacity refinement is the cleaner formulation for §3.6.3 / §3.6.4 / T-12 / T-13 in a v3.7.0 or later cleanup pass; T-9 makes both forms available so the eventual cleanup can be executed coherently rather than requiring all the citing locations to be repaired simultaneously.

Tuning of λ. In Form 3.2, the parameter λ trades off predictive gain against complexity cost. Empirically λ is observed to vary with affective state — high $|E(b)|$ (preprint Eq. T9-10) effectively raises λ at the component level, making affectively-marked components more pruning-resistant. This is the formal account of emotional memory enhancement (preprint §3.6.5, Pass III).

§4. Pruning Condition — Threshold Form

The pruning condition uses the threshold form rather than the strict-positivity form of main-paper Eq. T9-4. The OpenAI review of T-12 (appendix-corrections memo §2.8 Correction 3) correctly noted that the strict $I = 0$ condition for pruning is too brittle: real components have weak indirect predictive contributions even when their primary predictive role is excluded by filtered input.

The threshold-form pruning condition:

$$\text{Prune } \theta_i \quad \text{if} \quad G_i(t, \tau) \ < \ C_i \ - \ \epsilon \tag{T9.4-1}$$

with $\epsilon > 0$ a small *retention buffer* tuning the codec's pruning aggressiveness. Equivalent inequality forms:

$$G_i(t, \tau) - C_i \ < \ -\epsilon \quad \Longleftrightarrow \quad I(\theta_i; X_{t+1:t+\tau} \mid \theta_{-i}) \ < \ C_i - \epsilon \tag{T9.4-2}$$

Comparison to main paper Eq. T9-4. The main paper writes $\Delta_{\text{MDL}}(\theta_i) < 0$ as the pruning trigger, which corresponds to $\epsilon = 0$ — strict break-even. T-9 generalises by introducing the retention buffer ϵ, which more accurately models biological pruning dynamics (where small predictive contributions are preserved against transient noise) and synthetic-codec pruning hyperparameters (where threshold-based deletion is standard).

The strict break-even case is recovered as $\epsilon \to 0$, so the T-9 form does not invalidate the existing T9-4 citations in T-12 and T-13; it generalises them.

Implication for Narrative Drift (cross-reference T-12). Under filtered input $X' = \mathcal{F}(X)$ with the excluded signal $\mathcal{X}_{\text{excl}}$, components θ_i whose predictive contribution is exclusively to $\mathcal{X}_{\text{excl}}$ satisfy $G_i(t, \tau) \to 0$ on the filtered stream (because their target is absent from the observed input). The pruning condition (T9.4-1) then triggers because $0 < C_i - \epsilon$ for any positive cost component. T-12 Theorem T-12's irreversibility result follows from this triggering plus the four-modality distinction in §5 below.

§5. Four Pruning Modalities

The pruning operation (T9.4-1) admits four distinct implementations in the codec, with different reversibility properties. The distinction matters for the recovery condition (§6) and for the Narrative-Drift irreversibility claim in T-12 Correction 1 (appendix-corrections memo §2.8).

Modality 5.1 — Reversible suppression. The component θ_i's output weighting is reduced to zero (or below a participation threshold) but the component's parameters and structure remain stored in the codec. Recovery is straightforward: re-weighting restores the component. This is the operation underlying behavioural extinction in conditioning (the conditioned response weakens but the trace persists) and dropout-style regularisation in neural networks.

Modality 5.2 — Weight decay. The component's parameters decay continuously toward a default state under a regularisation pressure $\propto \lambda$. The component is not deleted but loses fidelity; partial recovery is possible if the default state is informative.

Modality 5.3 — Representational forgetting. The component's parameters are overwritten by competing components during consolidation (Pass II, preprint §3.6.4). The structural slot persists but the specific representation is lost. Recovery requires re-exposure to the relevant input stream during a subsequent Maintenance Cycle and is partial (the re-learned representation differs from the original in fine-grained detail).

Modality 5.4 — Architectural pruning. The component's parameters and structural slot are both deleted; the codec architecture is reduced. Recovery is impossible at the codec level — the component must be regrown from scratch through a full learning episode. This is the irreversible modality.

Modality classification under filtered input. T-12 Theorem T-12's "irreversibility" claim (as stated in the existing preprint) requires Modality 5.4 (architectural pruning) and excludes Modalities 5.1–5.3. T-9 makes this modality-dependence explicit; the v0.4 appendix-corrections memo §2.8 Correction 1 ("irreversible should be conditional on no protected archive / no replay buffer / no external teacher / no architectural reserve capacity / continued operation under the same filter / pruning is literal capacity deletion, not reversible suppression") aligns with the Modality 5.4 reading.

Real biological and synthetic codecs typically exhibit a *mix* of modalities, with Modality 5.4 reserved for components persistently pruned across many Maintenance Cycles. The transition from reversible to irreversible pruning under sustained filtered input is the structural mechanism underlying chronic Narrative Drift (T-12).

§6. Recovery Condition

A pruned component θ_i is *recoverable* if there exists a process by which it can be restored to active participation in the codec. The recovery probability over a recovery window τ_R is:

$$P\big(\text{recover } \theta_i \mid \tau_R\big) \;=\; P\big(\text{Modality 5.1 or 5.2}\big)\cdot p_{\text{restore}}(\tau_R) + P\big(\text{Modality 5.3 or 5.4}\big)\cdot p_{\text{regrow}}(\tau_R) \tag{T9.6-1}$$

The first term covers reversible / partially reversible pruning (suppression, weight decay); the second covers representational forgetting and architectural pruning, where recovery requires external input.

Recovery is positive only if at least one of three conditions holds:

1. **Protected memory.** The codec retains an archived representation of θ_i in a non-pruned substrate (separate cache, version-controlled backup,

neuro-physiologically protected memory consolidated to a different region). Modalities 5.1 and 5.3 can recover under this condition.

2. **External teacher / re-exposure.** The codec is exposed to input streams containing the signal $\mathcal{X}_{\text{excl}}$ that the pruned component was originally tracking. Active relearning during a subsequent Maintenance Cycle Pass II rebuilds the component (with caveats about fine-grained fidelity). All four modalities can recover under this condition over sufficient time, though Modality 5.4 requires a full learning episode comparable to the original acquisition.

3. **Architectural reserve.** The codec has structural slots that were not committed to specific components and can be allocated to host the regrown representation. This is the condition under which Modality 5.4 recovery is mechanically possible at all.

If none of (1), (2), (3) holds, then $P(\text{recover}\,\theta_i \mid \tau_R) = 0$ for all τ_R, and the pruning is permanent.

Substrate Fidelity Condition. T-12's Substrate Fidelity Condition (Theorem T-12b — redundancy of δ-independent input channels crossing the Markov blanket) is the lineage-scale analogue of (2): the channels ensure that the input stream continues to contain the substrate-relevant signal even under filtering by external mechanisms $\mathcal{F}$. T-9's recovery condition supplies the within-codec implementation: protected components, replay buffers, architectural reserve.

§7. Corollaries — Narrative Drift and Action-Drift

T-9's primitives support two corollary chains developed in appendices T-12 and T-13.

Corollary 7.1 — Narrative Drift (T-12). Under sustained filtered input $X' = \mathcal{F}(X)$ excluding signal $\mathcal{X}_{\text{excl}}$: - Components θ_i whose predictive gain is exclusively on $\mathcal{X}_{\text{excl}}$ have $G_i(t,\tau) \to 0$ on the filtered stream. - The pruning condition (T9.4-1) triggers across all such components. - If the pruning is in Modality 5.4 (architectural) — which dominates under sustained filtering across many Maintenance Cycles — and none of the recovery conditions (§6 items 1–3) holds, the capacity to model $\mathcal{X}_{\text{excl}}$ is permanently lost. - The codec cannot detect its own capacity loss from within (the lost components no longer participate in prediction error generation), reproducing T-12a's non-identifiability claim.

The full formal treatment is in T-12; T-9 supplies the modality-specific reading of "irreversible" that T-12 Correction 1 requires.

Corollary 7.2 — Action-Drift (T-13). Components encoding behavioural-evaluation capacity for unused branches: - Have predictive gain $G_i(t,\tau)$ measured against the input stream's actually-realised branch outcomes; if certain branches are never selected, the evaluators have no training signal. - The pruning condition triggers when the unused evaluator's G_i falls below $C_i - \epsilon$. - Under Modality 5.4, the evaluator is permanently pruned; the codec becomes confidently impotent in the corresponding action domain.

T-13's Proposition T-13.P1 (Action-Drift) is the lineage-scale (behavioural-repertoire) instance of this within-codec mechanism.

Cross-reference: lineage-level Maintenance Cycle. Appendix T-15 §3 develops the structural correspondence between the within-life Maintenance Cycle and phylogenetic refinement. T-9's four pruning modalities map respectively to: temporary niche reduction (5.1), lineage drift under relaxed selection (5.2), niche replacement (5.3), and lineage extinction (5.4). The recovery conditions (§6) map to phylogenetic redundancy: protected refugia (1), ecological re-exposure under niche restoration (2), and developmental reserve capacity (3).

§8. Relationship to Main Paper §3.6 Equations

T-9 is consolidating, not displacing. The main-paper equations T9-1 through T9-13 (preprint §3.6.1–§3.6.6) are preserved as cited; T-9 introduces additional formal primitives and refinements that supplement them.

Main paper	T-9
T9-1 ($K(P_\theta(t)) \leq C_{\text{ceil}}$) — total complexity ceiling	§1 setup
T9-2 ($\mathcal{M}_\tau : P_\theta(t) \to P_\theta(t+\tau)$) — Maintenance Cycle operator	§1 setup
T9-3 ($\Delta_{\text{MDL}}(\theta_i) = I(\theta_i; X \mid \theta_{-i}) - \lambda K(\theta_i)$) — MDL pruning quantity	§2 predictive gain G_i + §3 maintenance cost C_i (Form 3.2 K-approximation)
T9-4 (Prune if $\Delta_{\text{MDL}} < 0$) — pruning condition	§4 threshold form (T9.4-1 with $\epsilon \to 0$)
T9-5 (Landauer pruning cost) — thermodynamic floor	§5 modality dependence (irreversibility applies to Modality 5.4)
T9-6 (ΔK_{prune}) — pruning capacity recovery	§3 + §5 (resource-capacity form makes the bookkeeping additive over modalities)
T9-7 / T9-8 ($\Delta K_{\text{compress}}$) — consolidation gain	§1 setup (Pass II) — T-9 does not re-derive consolidation
T9-9 / T9-10 ($w(b)$, $E(b)$) — REM importance weighting	§3 (affective tuning of λ) — T-9 does not re-derive REM sampling
T9-11 — REM sampling distribution	unchanged — T-9 does not re-derive Pass III
T9-12 / T9-13 — net complexity budget	§1 setup — T-9's resource-capacity form refines the budget bookkeeping

Net new content in T-9: explicit predictive gain $G_i(t, \tau)$ definition (§2); resource-capacity cost framing as primary (§3 Form 3.1); threshold-form pruning condition with retention buffer ϵ (§4); four pruning modalities (§5); recovery condition (§6); modality-specific reading of T-12's irreversibility claim (§7.1).

§9. Open Edges

Coordination with T-12 channel-independence reformulation (Phase 4). T-12 is in the appendix-corrections queue (v0.4 §2.8) for a reformulation of the channel-independence condition: independence of *filtering mechanisms*, not signals. T-9's pruning condition (§4) and recovery condition (§6) are written to coordinate with that reformulation, but T-12's Theorem T-12 proof will need to be re-examined once the reformulated channel-independence definition is in place. Specifically: the irreversibility claim in T-12 §3.1 currently cites T9-3 / T9-4; under the v3.7.0 cleanup it should cite T-9's §4 threshold form + §5 modality classification + §6 recovery condition, with the irreversibility reading restricted to Modality 5.4 under the no-recovery-condition case. Open.

Resource-capacity vs. K-complexity bookkeeping reconciliation. §3 makes both forms available but does not derive their quantitative correspondence. For some component classes the two are closely related ($C_i^{\text{params}} \sim K(\theta_i)$ within a constant factor for memorised lookup tables, for example); for others they diverge sharply (compositional structure shared across components has K-complexity savings the resource-capacity form does not capture). A v3.7.0 or later reconciliation is desirable. Open.

Virtual-reading neutrality (v3.6.21). The fully-virtual standing-state reading (main paper §8.6.1) re-describes the Maintenance Cycle as properties of the filter-passing stream rather than a running machine, but does **not** re-tier the Form 3.1 / Form 3.2 bookkeeping: Form 3.1 (resource-capacity) remains primary for all operational claims, and T-12's operative proof continues to use it. The stream-native compressibility reading enters only as the interpretive layer noted in T-12 §3.1. The K-additivity reconciliation above is the place any future Form re-tiering would be argued — not the virtual reading. Open (do not conflate with the v3.7.0 cleanup).

Empirical calibration of ϵ. The retention buffer ϵ in (T9.4-1) is an effective pruning hyperparameter. Empirical biological values would come from neural-pruning studies (synaptic decay thresholds, dendritic-spine retention rates) or from the Δ_self^op asymptote experiment in the `opt-ai-subject` prototype. T-9 does not derive a specific value. Open.

Cross-link to Maintenance Cycle empirical predictions. Preprint §3.6.7 lists empirical predictions for the Maintenance Cycle (sleep / dream / consolidation). T-9's four pruning modalities make finer-grained predictions: the prediction that "REM dreams disproportionately sample high-importance branches" (preprint §3.6.5, Pass III) decomposes into modality-specific predictions about which kinds of representations are preserved by Modality 5.1 (importance-weighted retention against pruning) vs. Modality 5.4 (where the absence of high-importance branches in waking experience leads to architectural deletion of the corresponding evaluator). Open.

This appendix is maintained as part of the OPT project repository alongside opt-theory.md. References to the Maintenance Cycle primitives in preprint §3.6 are preserved; T-9 supplements with explicit predictive-gain G_i (§2), resource-capacity cost (§3 Form 3.1), threshold-form pruning condition with retention buffer ϵ (§4), four pruning modalities (§5), and recovery conditions (§6). Corollary references: T-12 (Narrative Drift) §3.6.3; T-13 (Action-Drift) §6; T-15 (Phylogenetic Stability Filter) §3.