# CERTIFIED CONSTRAINT LIFECYCLES FOR TRAJECTORY-PROXY VIDEO DIFFUSION CONTROL

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Constraint timing poses a dilemma in reverse denoising: later updates can violate previously satisfied constraints, while premature protection can hinder progress on unmet ones. To address this tension, we introduce Certified Constraint Lifecycles (CCL). CCL formulates inference-time diffusion control as an admission–preservation lifecycle under evolving protection obligations: qualifying observer-supported satisfaction on realized post-update states triggers irreversible gate-level admission, and the accumulated commitment state determines which protection relations and nonlinear endpoint checks apply to subsequent denoising updates. For differentiable trajectory proxies, CCL projects the sampler proposal onto the one-sided first-order cone induced by the exposed Jacobian and then applies a finite-step nonlinear guard. For successfully completed transitions, we prove guarded commitment invariance: every admitted proxy commitment remains satisfied. At a fixed state and proposal, with the exposed Jacobian evaluated exactly, we derive an exact pre-guard decomposition of the feasible row-space motion retained by one-sided protection but removed by equality protection. Controlled experiments show that CCL more reliably achieves and maintains trajectory-proxy satisfaction while reducing collateral deviation under matched final proxy satisfaction.

## 1 INTRODUCTION

Diffusion models provide powerful generative priors for image and video synthesis (Sohl-Dickstein et al., 2015; Ho et al., 2020; Rombach et al., 2022; Karras et al., 2022; Blattmann et al., 2023), and their denoising trajectories can be steered through a variety of control mechanisms. For differentiable constraints, however, a natural inference-time formulation minimizes a residual at the current denoising state or an objective on the final sample. This treats satisfaction as state-local: satisfying a constraint does not itself change what subsequent updates must preserve. A constraint satisfied at one step may therefore be violated during later refinement, whereas premature protection can unnecessarily restrict subsequent acquisition. This raises two coupled questions: when should realized satisfaction become a persistent commitment, and how should later updates preserve that commitment while remaining close to the pretrained sampler's proposal?

We introduce Certified Constraint Lifecycles (CCL). CCL formulates inference-time diffusion control as an admission–preservation lifecycle under evolving protection obligations: qualifying observer-supported satisfaction on realized post-update states triggers irreversible gate-level admission, and the accumulated commitment state determines which protection relations and nonlinear endpoint checks apply to subsequent denoising updates. It admits a constraint only when its declared proxy criterion has held at every realized post-update state during the prescribed dwell and its state-dependent prerequisites are satisfied. The committed set evolves along the realized denoising trajectory rather than being fixed in advance. As commitments form, the set of exposed protection relations evolves: a structured relation is exposed only when all constraints in its support are committed (Figure 1).

Given the current commitments, CCL protects only the structured relations exposed by that commitment state. For differentiable structured relations, CCL projects the sampler proposal onto the exposed first-order cone to obtain the feasible update closest to the proposal in Euclidean distance. It then applies a finite-step nonlinear guard before execution. For successfully completed transitions, we prove guarded commitment invariance relative to the declared observer: every admitted proxy commitment remains satisfied. For the same state and proposal, with the exposed Jacobian evaluated



Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

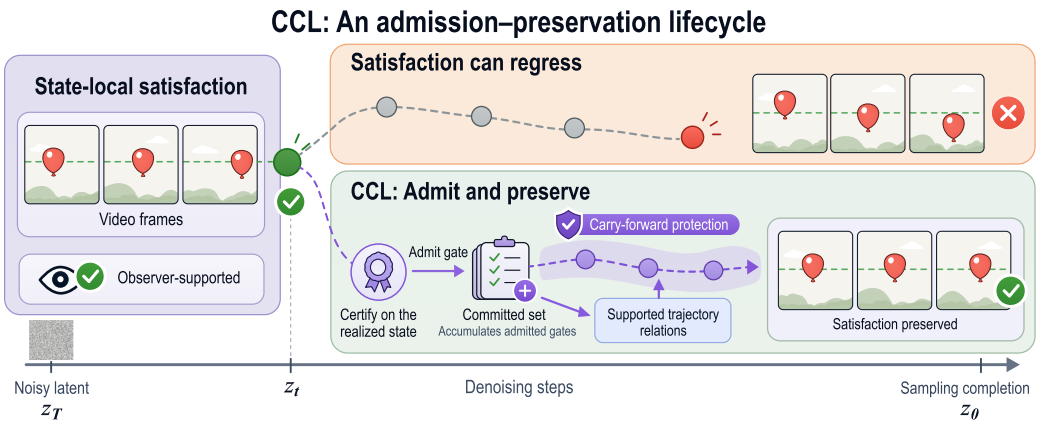


Figure 1: The figure contrasts state-local proxy satisfaction, which can regress, with qualifying post-update satisfaction admitted as a persistent commitment. The resulting committed set determines which trajectory relations are exposed and governs carry-forward protection during denoising.

exactly, we derive an exact pre-guard decomposition of the feasible row-space motion retained by one-sided protection but removed by equality protection.

We evaluate lifecycle behavior and collateral deviation in separate protocols. A preregistered staggered-admission study separates certification from carry-forward protection, while a projection–guard factorial identifies how the two components contribute to progress and persistence. We connect mechanism to rollout behavior by comparing acquisition across protection geometries and quantifying collateral deviation under matched final proxy satisfaction. On 100 paired CogVideoX-2b (Yang et al., 2024) seeds with matched final proxy satisfaction, CCL reduces deviation from the corresponding aligned-free sample relative to scalar rescaling, with median per-seed reductions of 82.21%–89.38% across pixel, latent, LPIPS (Zhang et al., 2018), and DINO (Caron et al., 2021) distances. Under the same commitment lifecycle, one-sided protection yields lower final-output deviation than equality protection, with median per-seed reductions of 20.86% to 29.64% across the four measures. CCL also improves acquisition and sustained adherence to externally specified object trajectories, with success assessed independently of the control observer.

We make three contributions. First, we introduce an admission–preservation lifecycle that turns qualifying post-update satisfaction into persistent commitments whose accumulation governs subsequent protection. Second, we prove guarded commitment invariance for successfully completed transitions and derive an exact pre-guard decomposition that quantifies the reduction in proposal deviation relative to equality protection. Third, we show that CCL improves acquisition and sustained adherence to externally specified object trajectories and reduces collateral deviation relative to scalar rescaling and equality protection under matched final proxy satisfaction.

## 2 Related Work

### 2.1 Inference-Time Control of Diffusion Models

Inference-time diffusion control introduces specified conditions into the reverse process through several mechanisms. Classifier and classifier-free guidance modify the denoiser prediction through external gradients or conditional–unconditional mixing (Dhariwal & Nichol, 2021; Ho & Salimans, 2021), with manifold-aware variants reducing off-manifold deviation (Chung et al., 2025). Posterior-sampling and loss-guided methods derive corrections from measurement consistency or task objectives (Chung et al., 2023; Bansal et al., 2024; Yu et al., 2023; He et al., 2024; Xie et al., 2023), while structured controllers inject spatial conditions through branches, adapters, or attention mechanisms (Zhang et al., 2023; Li et al., 2023; Mou et al., 2024). These methods specify how conditions influence denoising. CCL takes a proxy interface as given and studies how the evolving commitment state changes subsequent protection obligations.

### 2.2 Trajectory and Motion Control in Video Generation

Video generation methods control motion through structured conditioning signals, including spatiotemporal cues, sparse object trajectories, and explicit camera paths. VideoComposer combines spatial and temporal controls, while DragNUWA directly conditions generation on object trajectories (Wang et al., 2023a; Yin et al., 2023). MotionCtrl separates camera and object motion through dedicated control pathways, and MagicMotion develops dense-to-sparse trajectory representations for fine-grained motion guidance (Wang et al., 2023b; Li et al., 2025). Tora encodes trajectories into motion features that condition a video diffusion transformer (Zhang et al., 2025b). CCL complements trajectory conditioning by turning qualifying post-update satisfaction into commitments that subsequent denoising updates must preserve.

### 2.3 Constrained Sampling and Persistent Commitments

Constrained diffusion draws on projection, barrier, and mirror techniques from constrained optimization (Nocedal & Wright, 2006; Boyd & Vandenberghe, 2004; Bertsekas, 1999). Reflected and mirror dynamics constrain sampling to prescribed domains, while projection- and trust-based methods repeatedly correct the current iterate or clean-sample estimate (Lou & Ermon, 2023; Fishman et al., 2023; Liu et al., 2023; Christopher et al., 2024; Huang et al., 2024a). More recent formulations incorporate constrained Langevin updates, posterior-mean projection, or adaptive enforcement schedules (Zhang et al., 2025a; Narasimhan et al., 2025; Trupin & Xue, 2026; He et al., 2026). SafeDiffuser establishes finite-time diffusion invariance for prespecified safety constraints and introduces relaxed and time-varying enforcement to mitigate local traps caused by imposing those constraints early in denoising (Xiao et al., 2025). CCL instead admits a gate as a persistent commitment only when realized post-update states provide qualifying observer-supported evidence of its satisfaction.

LTLDoG uses a differentiable satisfaction measure for a prespecified temporal-logic formula to guide each reverse step (Feng et al., 2024), while hint$^2$ uses an automaton state updated from environment execution history to guide subsequent action-chunk sampling (Zoellner et al., 2026). Adaptive correction scheduling uses one-step constraint defect to allocate a limited correction budget across a generative rollout (Trupin & Xue, 2026). Conventional active-set methods update a working set while solving a fixed optimization problem, whereas CCL accumulates gate-level commitments through post-update admissions within a single reverse-denoising rollout. The committed set determines the exposed protection relations and applicable nonlinear endpoint checks, while the current denoising state determines the relations' local geometry.

## 3 Certified Constraint Lifecycles

CCL separates commitment admission from commitment preservation. The accumulated commitments determine which protection relations and nonlinear endpoint checks apply to the next update. This update must preserve commitments already in force before new constraints can be admitted on the realized post-update state. When the admission conditions are satisfied, observer-supported proxy satisfaction becomes a persistent commitment, extending the set that governs subsequent updates (Figure 2).

### 3.1 Commitment State and Post-update Admission

A scalar proxy gate indicates whether a specified proxy criterion holds at a given state; the commitment state separately encodes whether the corresponding constraint has been admitted and must remain satisfied. CCL bases admission on observer-supported gate satisfaction at realized post-update states. For a state $z$ and an optional fixed reference $r$, the frozen observer returns a measurement $m(z; r)$ and a binary flag $\chi(z; r)$ indicating whether the measurement is supported. We assume a finite family of scalar proxy gates, each satisfied when

$$\chi(z; r_i) = 1 \quad \text{and} \quad c_i(z) \le \epsilon_i, \qquad i \in \mathcal{P}, \tag{1}$$

where $c_i(z)$ is the residual induced by the observer measurement $m(z; r_i)$ and $\epsilon_i$ is its prescribed threshold.

Gate satisfaction alone does not determine admission, which may also depend on earlier commitments or other declared state conditions. We collect previously admitted constraints in the committed set

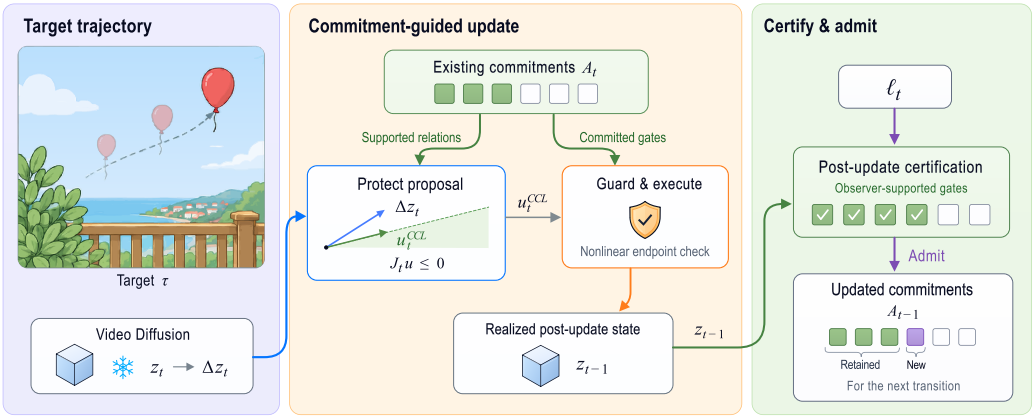


Figure 2: Certified Constraint Lifecycles. CCL projects the proposed diffusion update and applies nonlinear endpoint checks to protect existing commitments. Qualifying gates are certified on the realized post-update state and admitted as commitments for subsequent protection.

$A_t \subseteq \mathcal{P}$. CCL admits an uncommitted constraint when it is currently eligible and its gate has been satisfied at each of the last $K$ realized post-update states. New admissions can make other constraints eligible, so CCL resolves eligibility dependencies in a fixed topological order. Each admission updates the committed set before the next constraint is evaluated. The resulting set carries all admitted constraints into subsequent denoising steps.

### 3.2 Lifecycle-conditioned protection geometry

In the projection realization, a committed proxy gate specifies what must remain satisfied, while the commitment state determines which local relations constrain the next update. Applying all protection relations from the outset can make uncommitted constraints harder to satisfy. CCL therefore exposes a relation only when all constraints in its declared support have been admitted.

The commitment state selects the exposed relations, while the current denoising state determines their local orientations and Jacobians. Protecting an exposed relation does not require holding its value fixed. CCL orients each exposed relation so that a positive directional derivative corresponds to increasing relation magnitude:

$$\sigma_j(t) = \begin{cases} +1, & h_j(z_t) \geq 0, \\ -1, & h_j(z_t) < 0, \end{cases} \qquad J_j(z_t) = \sigma_j(t) \nabla h_j(z_t)^\top. \tag{2}$$

Here $h_j$ denotes a differentiable signed relation. For a nonzero relation, one-sided protection prevents first-order increases in magnitude but, unlike equality protection, still allows decreases. Among updates that satisfy these constraints, CCL selects the one closest to the sampler proposal $\Delta z_t$:

$$u_t^{\mathrm{CCL}} = \arg\min_u \frac{1}{2}\|u - \Delta z_t\|_2^2 \quad \text{subject to} \quad J_t u \leq 0, \tag{3}$$

where $J_t$ collects the oriented gradients of the exposed relations as rows. The projection leaves any proposal that already satisfies these constraints unchanged.

For trajectory control, reducing position errors at individual frames need not reduce their temporal differences. We therefore include first- and second-order temporal differences alongside position error in the protection relations. Let $q_k(z)$ denote the observer's differentiable position readout for frame $k$. For the prescribed target $\tau_k$, the position error and scalar gate residual are

$$e_k(z) = q_k(z) - \tau_k, \qquad c_k(z) = \|e_k(z)\|_2. \tag{4}$$

The residual measures distance to the target, so the gate threshold $\epsilon_k$ specifies the allowed positional error. Appendix B.1 specifies the normalized relations and their commitment supports.

### 3.3 GUARDED CCL TRANSITION

The projection enforces local first-order conditions, but a finite step can still violate a committed gate. Preserving existing commitments requires both observer support and proxy satisfaction at the endpoint. For a committed set $A$, the endpoints satisfying these requirements form the feasible set

$$\mathcal{F}(A) = \{z : \chi(z; r_i) = 1,\ c_i(z) \le \epsilon_i,\ \forall i \in A\}. \tag{5}$$

Nonnegative scaling preserves the projected update's first-order feasibility. Given $z_t \in \mathcal{F}(A_t)$ and a finite set of candidate step sizes $\Gamma \subset (0, 1]$, CCL selects the largest feasible candidate step to retain as much of the projected update as possible:

$$\begin{aligned} \alpha_t^\star &= \max\left(\left\{\alpha \in \Gamma : z_t + \alpha u_t^{\mathrm{CCL}} \in \mathcal{F}(A_t)\right\} \cup \{0\}\right), \\ z_{t-1} &= z_t + \alpha_t^\star u_t^{\mathrm{CCL}}. \end{aligned} \tag{6}$$

The guard preserves existing commitments, while new constraints are admitted only at the realized post-update state $z_{t-1}$.

The same endpoint guard applies to complete action candidates when linear scaling would alter their structure. CCL selects the first feasible candidate in a prescribed order and rejects the transition if none is feasible. New constraints are then admitted from the realized state according to Section 3.1.

### 3.4 THEORETICAL PROPERTIES

The construction yields two complementary results. At the lifecycle level, admitted proxy commitments remain satisfied across successfully completed transitions, regardless of how candidates are constructed. At the geometric level, the one-sided projection minimizes deviation from the proposal over the exposed first-order cone and can retain feasible motion that equality protection removes.

**Theorem 1 (Guarded commitment invariance)** *Assume that $\mathcal{P}$ is finite, $z_t \in \mathcal{F}(A_t)$, and admission follows the irreversible rule of Section 3.1. Any successfully completed CCL transition satisfies*

$$A_t \subseteq A_{t-1}, \qquad z_{t-1} \in \mathcal{F}(A_{t-1}). \tag{7}$$

*Consequently, once a constraint is admitted, its observer-supported proxy gate remains satisfied after every subsequent successfully completed CCL transition.*

The guard preserves feasibility of the existing commitments, while post-update admission adds only constraints whose eligibility conditions and observer-supported proxy gates hold on the realized state $z_{t-1}$. Together with irreversible admission, these properties establish the lifecycle invariant independently of candidate construction.

**Corollary 2 (Exact decomposition relative to equality protection)** *At a fixed state, let $d = \Delta z_t$ be the proposal and $J_t$ the exact exposed Jacobian. Equality protection rules out all first-order changes in the exposed relations, including those allowed by the one-sided constraints. Equality protection therefore removes the entire row-space component of $d$, yielding $u_t^{\mathrm{eq}}$. One-sided protection instead projects this component onto the feasible part of the row space to obtain $q_t$, while leaving the null-space component unchanged:*

$$u_t^{\mathrm{CCL}} = u_t^{\mathrm{eq}} + q_t, \qquad \|u_t^{\mathrm{eq}} - d\|_2^2 - \|u_t^{\mathrm{CCL}} - d\|_2^2 = \|q_t\|_2^2. \tag{8}$$

Thus, retaining $q_t$ reduces squared deviation from the proposal by exactly $\|q_t\|_2^2$ relative to equality protection. This comparison holds pointwise before the finite-step guard, with proofs provided in Appendix A.

## 4 EXPERIMENTS

Our experiments examine two consequences of making realized satisfaction stateful: how commitments evolve along the denoising trajectory, and how much collateral deviation this protection introduces. A preregistered staggered-admission study first separates certification from carry-forward protection, while a projection–guard factorial identifies how the two components contribute to progress and persistence. We then connect mechanism to rollout behavior by comparing acquisition across protection geometries and quantifying collateral deviation under matched final proxy satisfaction. Finally, an object-trajectory study evaluates acquisition and sustained adherence to externally specified targets, with object-level success assessed independently of the control observer.

Table 1: Staggered-admission outcomes on 32 paired seeds. Targets are derived from aligned-free readouts and screened for regression opportunities (Appendix C.1). Completion requires all gates to hold at the final state.

| Method | All-gate completion | Seeds with regression | Regression cell–steps | Zero steps |
|---|---|---|---|---|
| Certify-only | 0/32 | 32/32 | 7,247 | 0 |
| Projection-only | 8/32 | 25/32 | 178 | 0 |
| Guard-only | 0/32 | 0/32 | 0 | 576 |
| Static-All | 17/32 | 0/32 | 0 | 0 |
| **CCL** | **30/32** | 0/32 | 0 | 0 |

### 4.1 From Satisfaction to Persistent Commitment

Carry-forward protection changes the lifecycle outcome under staggered admission. CCL completes 30/32 trajectories, whereas Certify-only completes none (Table 1). CCL exhibits no post-admission regression, while Certify-only accumulates 7,247 regression cell–steps after first certification. The two remaining CCL failures reflect incomplete acquisition. A controlled late-conflict example illustrates the effect in decoded video: after admission, the unprotected branch departs from the target trajectory, while CCL and CCL–ExactSym preserve the prior commitment (Figure 3).

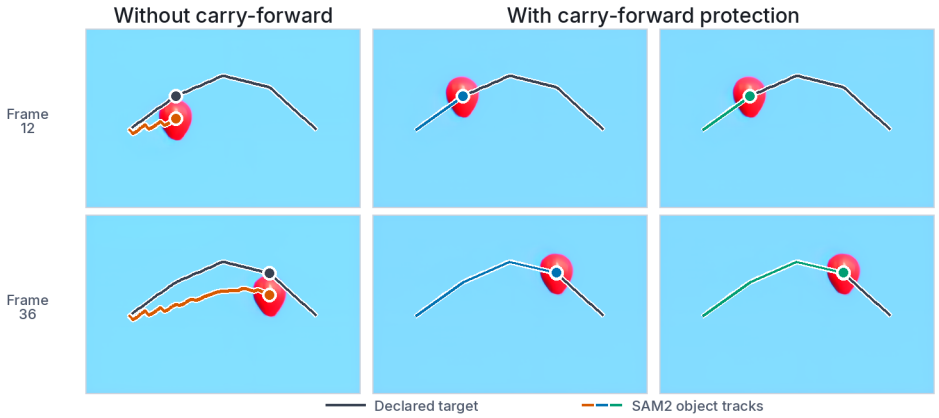


Figure 3: **Carry-forward protection preserves an admitted trajectory commitment.** Under the same controlled late conflict, the unprotected branch departs from the declared trajectory, whereas the protected branches preserve the prior commitment through subsequent denoising.

Acquisition also depends on when protection relations are exposed. Static-All differs from CCL only in exposing the full relation family from the first controlled transition. It completes 17/32 trajectories, compared with 30/32 for CCL. Both methods nevertheless preserve their admitted commitments without taking a zero step. This pattern suggests that exposing protection relations before their supporting commitments are admitted can impede subsequent acquisition.

Ablating projection and the guard separately reveals complementary failure modes. Projection-only succeeds in 8/32 runs but permits post-admission regression in 25/32. Guard-only avoids regression but frequently stalls progress: it takes a zero step in 576 of 960 transitions, and no run meets the final criterion. With both components, CCL succeeds in 30/32 runs without regression or zero steps. Projection thus supports progress toward full acquisition, while the guard keeps subsequent updates from undoing admitted commitments.

Beyond the component ablations, we assess whether CCL improves completion over alternative ways of enforcing the same constraints. In a separate held-out comparison, CCL satisfies all gates at the final three states in 60 of 64 runs, compared with 46 for Static-All Restoration and 44 for DirectGate-Jacobian. Both comparisons remain significant after Holm correction (Holm, 1979) (Table 2).

Table 2: Held-out comparison of constraint handlers on 64 paired seeds. Success requires all 32 gates to hold at the final three states. Differences are CCL minus comparator, in percentage points.

| Method | Terminal-window success | Difference (pp) | Nominal 95% CI (pp) | Holm-adjusted $p$ |
|---|---|---|---|---|
| **CCL** | **60/64** | – | – | – |
| Static-All Restoration | 46/64 | $+21.875$ | $[9.054, 34.961]$ | $0.00258$ |
| DirectGate-Jacobian | 44/64 | $+25.000$ | $[11.734, 38.283]$ | $0.00171$ |
| TR-SQP | 2/64 | $+90.625$ | $[79.831, 95.632]$ | $2.08 \times 10^{-17}$ |
| Guard-only Continuous | 0/64 | $+93.750$ | $[82.779, 97.543]$ | $8.67 \times 10^{-18}$ |
| Reversible-Dwell-Hysteresis | 0/64 | $+93.750$ | $[82.779, 97.543]$ | $8.67 \times 10^{-18}$ |

Table 3: Collateral deviation under matched final satisfaction across two target families. Mean differences $\Delta = D_{\text{comp}} - D_{\text{CCL}}$ are reported with 95% paired-bootstrap CIs; reductions are median per-seed relative reductions.

| Metric | Scalar $\Delta$ | Reduction | CCL–ExactSym $\Delta$ | Reduction |
|---|---|---|---|---|
| **Primary study (100 paired seeds)** | | | | |
| Pixel | 0.0673 [0.0647, 0.0695] | 82.21% | 0.0067 [0.0052, 0.0082] | 24.73% |
| Latent | 81.688 [78.977, 84.074] | 86.83% | 3.881 [3.007, 4.755] | 20.86% |
| LPIPS | 0.3460 [0.3356, 0.3553] | 88.56% | 0.0291 [0.0232, 0.0362] | 29.64% |
| DINO | 0.2483 [0.2402, 0.2557] | 89.38% | 0.0191 [0.0146, 0.0245] | 27.91% |
| **Parabolic replication (60 paired seeds)** | | | | |
| Pixel | 0.0667 [0.0632, 0.0698] | 82.56% | 0.0054 [0.0041, 0.0067] | 21.45% |
| Latent | 81.951 [77.765, 85.445] | 86.96% | 4.683 [3.596, 5.807] | 22.16% |
| LPIPS | 0.3462 [0.3295, 0.3598] | 89.50% | 0.0213 [0.0157, 0.0272] | 27.21% |
| DINO | 0.2501 [0.2380, 0.2608] | 89.98% | 0.0118 [0.0083, 0.0157] | 22.00% |

Irreversible admission raises a natural concern: a transient error in the admission readout may trigger a lasting commitment. We therefore examine whether requiring satisfaction at two consecutive states prevents such false admissions. A one-step readout perturbation causes false admissions on 4 of 64 paired tasks with FirstHit, but none with consecutive confirmation (Appendix D.9).

### 4.2 Compatible Motion and Collateral Deviation

Preserving an admitted commitment need not eliminate motion that remains compatible with it. The pre-guard decomposition in Corollary 2 identifies the feasible row-space motion retained by one-sided protection and removed by equality protection. We compare acquisition across protection geometries under the same lifecycle rules and nonlinear guard. Including temporal relations in one-sided protection improves acquisition. On a separate set of 32 paired held-out seeds, structured one-sided protection completes 30 trajectories, compared with 21 for gate-radial protection and 22 for position-only protection. The benefit of the structured family also depends on how its relations are enforced. Replacing one-sided protection with equality protection on the same family reduces completion to 3/32 (Appendix Table 8).

The matched-satisfaction study next compares collateral deviation after complete denoising rollouts, using targets that are satisfied at initialization. CCL, scalar full-gate bisection, and CCL–ExactSym satisfy the common final proxy criterion on all 100 paired CogVideoX-2b seeds. Deviation is measured relative to the corresponding aligned-free sample generated under the same calibrated schedule. Against scalar bisection, which preserves proposal direction, CCL has lower deviation on every seed under all four metrics: pixel and latent distances, LPIPS, and DINO feature distance. Median per-seed reductions range from 82.21% to 89.38% (Table 3). When scalar bisection shortens a proposal, it also reduces compatible motion along with the rest of the update. Preserving proposal direction therefore does not guarantee lower deviation in the final generated sample.

The comparison with CCL–ExactSym isolates the effect of the protection geometry. CCL–ExactSym retains the same lifecycle rules and finite-step guard but replaces one-sided protection with equality protection. Median per-seed reductions range from 20.86% to 29.64% across the four metrics.




Figure 4: Comparison for a prescribed rightward trajectory. Tora+CCL (bottom) moves the bus in the target direction, whereas Tora (top) generates motion in the opposite direction.

Table 4: Object-trajectory outcomes on 64 paired tasks. Primary success requires timely full acquisition and sustained satisfaction through sampling completion, with the same object intact and within positional tolerance in every frame.

| Method | Timely full acquisition | Primary success | Final-video success |
|---|---|---|---|
| Tora | 38/64 | 25/64 | 29/64 |
| All-frame | 43/64 | 32/64 | 36/64 |
| Tora+CCL | **49/64** | **45/64** | **47/64** |

Equality protection also suppresses compatible motion that improves the protected relations, yet offers no gain in final proxy success.

For parabolic targets, the median per-seed reductions against both comparators remain similar to those in the primary study (Table 3). All three methods again satisfy the final proxy criterion on all 60 paired seeds. The benefit of one-sided protection therefore extends to the parabolic target family. The advantage over scalar rescaling also extends to LTX-Video (HaCohen et al., 2024), with lower average deviation in both pixel and latent space on 32 held-out seeds (Appendix Table 17).

### 4.3 Object-Level Acquisition and Persistence

We next examine whether CCL improves acquisition and sustained adherence to externally specified object trajectories. Target trajectories are specified relative to the object's position in the reference image and fixed before generation. Tora (Zhang et al., 2025b) is evaluated with and without CCL on 64 paired tasks in 16 scenes, covering linear and curved paths. We assess the generated object trajectories using human-verified annotations independent of the control observer.

With CCL, Tora more often acquires the full target trajectory in time and maintains it until sampling is complete. Tora+CCL succeeds on 45 of 64 tasks, compared with 25 for Tora, a paired difference of 31.3 percentage points (Table 4). The rate of timely full acquisition increases from 38/64 for Tora to 49/64 for Tora+CCL. Among the 22 tasks on which only Tora+CCL succeeds, Tora fails to acquire the full target in time on 13 and regresses after acquisition on the remaining nine. CCL's benefit therefore extends beyond preserving an acquired trajectory to improving acquisition itself.

The gains extend to both linear and curved trajectories: Tora+CCL succeeds on 21 of 32 linear tasks and 24 of 32 curved tasks, compared with 10 and 15 for Tora, respectively. Figures 4 and 5 illustrate improved adherence to the prescribed motion: Tora+CCL follows the target direction, whereas Tora generates motion in the opposite direction.

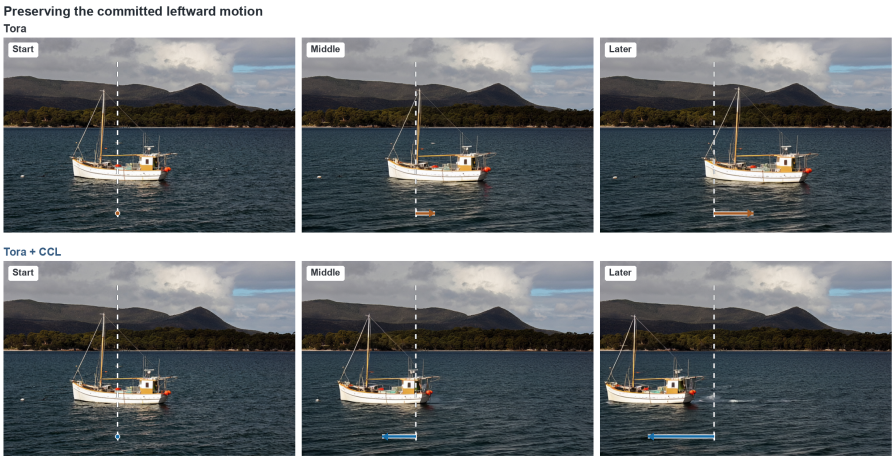


Figure 5: Comparison for a prescribed leftward trajectory. Tora+CCL (bottom) moves the boat in the target direction, whereas Tora (top) generates motion in the opposite direction.

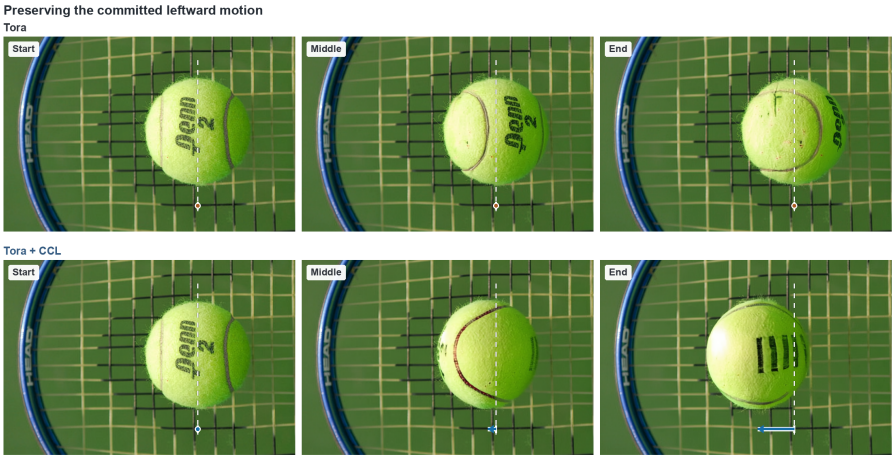


Figure 6: Comparison for a prescribed leftward trajectory. Tora+CCL (bottom) moves the tennis ball leftward, whereas Tora (top) shows little horizontal displacement.

Extending projection constraints to all frames lowers success on these tasks. Under the same admission and guard rules, the All-frame variant succeeds on 32 of 64 tasks, compared with 45 for Tora+CCL (Table 4). All-frame loses satisfaction after acquisition on eight tasks where Tora+CCL succeeds; the reverse occurs on one task. All-frame's broader projection constraints do not translate into better preservation after acquisition. This comparison supports using the commitment state to select projection constraints. Full results and computational costs are reported in Appendix G.

## 5 Conclusion

We introduced Certified Constraint Lifecycles (CCL), which turns qualifying post-update satisfaction into persistent commitments that govern subsequent protection. This lifecycle preserves admitted proxy commitments across successfully completed transitions, while one-sided protection retains locally compatible proposal motion. Controlled experiments show that CCL improves acquisition and sustained adherence to externally specified object trajectories and reduces collateral deviation under matched final proxy satisfaction. More broadly, these results suggest inference-time constraint control can be organized around evolving protection obligations rather than treating satisfaction as a step-local event.

## AI USE STATEMENT

We used generative AI tools to assist with manuscript editing and translation and to identify relevant literature. The authors take responsibility for the final content of the paper, including its references and claims.

## REPRODUCIBILITY STATEMENT

Proofs and theoretical details are provided in Appendix A. The primary matched-satisfaction study uses a frozen CogVideoX-2b (Yang et al., 2024) backbone under the fixed configuration of Table 5 (Appendix B), with fixed seeds, declared gates, and shared calibration. Within each auxiliary protocol, paired arms share the same frozen backbone, exogenous inputs, and target metadata. Metric definitions and statistical aggregation rules are specified in Appendices C and G. Additional experiments and diagnostic analyses are documented in Appendices D–F.

## REFERENCES

Arpit Bansal, Hong-Min Chu, Avi Schwarzschild, Soumyadip Sengupta, Micah Goldblum, Jonas Geiping, and Tom Goldstein. Universal guidance for diffusion models. In *International Conference on Learning Representations*, 2024.

Dimitri P. Bertsekas. *Nonlinear Programming*. Athena Scientific, 2 edition, 1999.

Andreas Blattmann, Robin Rombach, Huan Ling, Tim Dockhorn, Seung Wook Kim, Sanja Fidler, and Karsten Kreis. Align your latents: High-resolution video synthesis with latent diffusion models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023.

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *IEEE/CVF International Conference on Computer Vision*, 2021.

Ho Kei Cheng, Seoung Wug Oh, Brian Price, Joon-Young Lee, and Alexander Schwing. Putting the object back into video object segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3151–3161, June 2024.

Jacob K. Christopher, Stephen Baek, and Ferdinando Fioretto. Constrained synthesis with projected diffusion models. In *Advances in Neural Information Processing Systems*, volume 37, 2024.

Hyungjin Chung, Jeongsol Kim, Michael T. McCann, Marc L. Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *International Conference on Learning Representations*, 2023.

Hyungjin Chung, Jeongsol Kim, Geon Yeong Park, Hyelin Nam, and Jong Chul Ye. CFG++: Manifold-constrained classifier free guidance for diffusion models. In *International Conference on Learning Representations*, 2025.

Prafulla Dhariwal and Alexander Quinn Nichol. Diffusion models beat gans on image synthesis. In *Advances in Neural Information Processing Systems*, 2021.

Zeyu Feng, Hao Luan, Pranav Goyal, and Harold Soh. LTLDoG: Satisfying temporally-extended symbolic constraints for safe diffusion-based planning. *IEEE Robotics and Automation Letters*, 9 (10):8571–8578, 2024. doi: 10.1109/LRA.2024.3443501.

Nic Fishman, Leo Klarner, Valentin De Bortoli, Emile Mathieu, and Michael Hutchinson. Diffusion models for constrained domains. *Transactions on Machine Learning Research*, 2023.

Yoav HaCohen, Nisan Chiprut, Benny Brazowski, Daniel Shalem, Dudu Moshe, Eitan Richardson, Eran Levin, Guy Shiran, Nir Zabari, Ori Gordon, Poriya Panet, Sapir Weissbuch, Victor Kulikov, Yaki Bitterman, Zeev Melumian, and Ofir Bibi. LTX-Video: Realtime video latent diffusion. *arXiv preprint arXiv:2501.00103*, 2024.

Yutong He, Naoki Murata, Chieh-Hsin Lai, Yuhta Takida, Toshimitsu Uesaka, Dongjun Kim, Wei-Hsiang Liao, Yuki Mitsufuji, J. Zico Kolter, Ruslan Salakhutdinov, and Stefano Ermon. Manifold preserving guided diffusion. In *International Conference on Learning Representations*, 2024.

Zhilin He, Bowei Li, Jianlin Dou, Yuner Zhang, and Changliu Liu. Motion planning with model-based diffusion via constraint optimization and adaptive scheduling. *arXiv preprint arXiv:2607.14455*, 2026.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. In *NeurIPS Workshop on Deep Generative Models and Downstream Applications*, 2021.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, 2020.

Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2):65–70, 1979.

William Huang, Yifeng Jiang, Tom Van Wouwe, and C. Karen Liu. Constrained diffusion with trust sampling. In *Advances in Neural Information Processing Systems*, volume 37, 2024a.

Ziqi Huang, Yinan He, Jiashuo Yu, Fan Zhang, Chenyang Si, Yuming Jiang, Yuanhan Zhang, Tianxing Wu, Qingyang Jin, Nattapol Chanpaisit, Yaohui Wang, Xinyuan Chen, Limin Wang, Dahua Lin, Yu Qiao, and Ziwei Liu. VBench: Comprehensive benchmark suite for video generative models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024b.

Nikita Karaev, Ignacio Rocco, Benjamin Graham, Natalia Neverova, Andrea Vedaldi, and Christian Rupprecht. CoTracker: It is better to track together. In *European Conference on Computer Vision*, 2024.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. In *Advances in Neural Information Processing Systems*, 2022.

Quanhao Li, Zhen Xing, Rui Wang, Hui Zhang, Qi Dai, and Zuxuan Wu. Magicmotion: Controllable video generation with dense-to-sparse trajectory guidance. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 12112–12123, 2025.

Yuheng Li, Haotian Liu, Qingyang Wu, Fangzhou Mu, Jianwei Yang, Jianfeng Gao, Chunyuan Li, and Yong Jae Lee. GLIGEN: Open-set grounded text-to-image generation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023.

Guan-Horng Liu, Tianrong Chen, Evangelos A. Theodorou, and Molei Tao. Mirror diffusion models for constrained and watermarked generation. In *Advances in Neural Information Processing Systems*, 2023.

Aaron Lou and Stefano Ermon. Reflected diffusion models. In *International Conference on Machine Learning*, 2023.

Chong Mou, Xintao Wang, Liangbin Xie, Yanze Wu, Jian Zhang, Zhongang Qi, and Ying Shan. T2I-Adapter: Learning adapters to dig out more controllable ability for text-to-image diffusion models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 4296–4304, 2024. doi: 10.1609/aaai.v38i5.28226.

Sai Shankar Narasimhan, Shubhankar Agarwal, Litu Rout, Sanjay Shakkottai, and Sandeep P. Chinchali. Constrained posterior sampling: Time series generation with hard constraints. In *Advances in Neural Information Processing Systems*, volume 38, 2025.

Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer, 2 edition, 2006.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, Mahmoud Assran, Nicolas Ballas, Wojciech Galuba, Russell Howes, Po-Yao Huang, Shang-Wen Li, Ishan Misra, Michael Rabbat, Vasu Sharma, Gabriel Synnaeve, Hu Xu, Hervé Jégou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. DINOv2: Learning robust visual features without supervision. *Transactions on Machine Learning Research*, 2024.

Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma, Haitham Khedr, Roman Rädle, Chloe Rolland, Laura Gustafson, Eric Mintun, Junting Pan, Kalyan Vasudev Alwala, Nicolas Carion, Chao-Yuan Wu, Ross Girshick, Piotr Dollár, and Christoph Feichtenhofer. SAM 2: Segment anything in images and videos. In *International Conference on Learning Representations*, 2025.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.

Jascha Sohl-Dickstein, Eric A. Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*, 2015.

B. Stellato, G. Banjac, P. Goulart, A. Bemporad, and S. Boyd. OSQP: an operator splitting solver for quadratic programs. *Mathematical Programming Computation*, 12(4):637–672, 2020. doi: 10.1007/s12532-020-00179-2.

Toshiro Tango. Equivalence test and confidence interval for the difference in proportions for the paired-sample design. *Statistics in Medicine*, 17(8):891–908, 1998.

Noah Trupin and Yexiang Xue. Enforcing constraints in generative sampling via adaptive correction scheduling. *arXiv preprint arXiv:2605.11214*, 2026.

Xiang Wang, Hangjie Yuan, Shiwei Zhang, Dayou Chen, Jiuniu Wang, Yingya Zhang, Yujun Shen, Deli Zhao, and Jingren Zhou. Videocomposer: Compositional video synthesis with motion controllability. *arXiv preprint arXiv:2306.02018*, 2023a.

Zhouxia Wang, Ziyang Yuan, Xintao Wang, Tianshui Chen, Menghan Xia, Ping Luo, and Ying Shan. Motionctrl: A unified and flexible motion controller for video generation. *arXiv preprint arXiv:2312.03641*, 2023b.

Wei Xiao, Tsun-Hsuan Wang, Chuang Gan, Ramin Hasani, Mathias Lechner, and Daniela Rus. SafeDiffuser: Safe planning with diffusion probabilistic models. In *International Conference on Learning Representations*, 2025.

Jinheng Xie, Yuexiang Li, Yawen Huang, Haozhe Liu, Wentian Zhang, Yefeng Zheng, and Mike Zheng Shou. BoxDiff: Text-to-image synthesis with training-free box-constrained diffusion. In *IEEE/CVF International Conference on Computer Vision*, 2023.

Zhuoyi Yang, Jiayan Teng, Wendi Zheng, Ming Ding, Shiyu Huang, Jiazheng Xu, Yuanming Yang, Wenyi Hong, Xiaohan Zhang, Guanyu Feng, et al. Cogvideox: Text-to-video diffusion models with an expert transformer. *arXiv preprint arXiv:2408.06072*, 2024.

Shengming Yin, Chenfei Wu, Jian Liang, Jie Shi, Houqiang Li, Gong Ming, and Nan Duan. DragNUWA: Fine-grained control in video generation by integrating text, image, and trajectory. *arXiv preprint arXiv:2308.08089*, 2023.

Jiwen Yu, Yinhuai Wang, Chen Zhao, Bernard Ghanem, and Jian Zhang. FreeDoM: Training-free energy-guided conditional diffusion model. In *IEEE/CVF International Conference on Computer Vision*, 2023.

Jichen Zhang, Liqun Zhao, Antonis Papachristodoulou, and Jack Umenberger. Constrained diffusers for safe planning and control. In *Advances in Neural Information Processing Systems*, volume 38, 2025a.

Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *IEEE/CVF International Conference on Computer Vision*, 2023.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018.

Zhenghao Zhang, Junchao Liao, Menghao Li, Zuozhuo Dai, Bingxue Qiu, Siyu Zhu, Long Qin, and Weizhi Wang. Tora: Trajectory-oriented diffusion transformer for video generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2063–2073, 2025b.

Moritz Zoellner, Anastasios Manganaris, Ahmed H. Qureshi, and Rohan Paleja. hint$^2$: Hierarchical world models for inference-time temporal logic guidance. *arXiv preprint arXiv:2608.13678*, 2026.

# A Proofs and Theoretical Details

## A.1 Proof of Theorem 1

Consider any successfully completed CCL transition. Irrespective of candidate construction, the finite-step guard accepts an endpoint only if

$$z_{t-1} \in \mathcal{F}(A_t).$$

Admission evidence is updated only after this state has been realized. Every newly admitted constraint $i$ has a supported gate that has held on the last $K$ realized post-update states. Since $K \geq 1$, the gate holds at $z_{t-1}$, and hence

$$\chi(z_{t-1}; r_i) = 1, \qquad c_i(z_{t-1}) \leq \epsilon_i$$

on the same realized state. Since admission is irreversible,

$$A_t \subseteq A_{t-1},$$

and every constraint in $A_{t-1}$, whether previously committed or newly admitted, satisfies its declared scalar gate at $z_{t-1}$. Therefore

$$z_{t-1} \in \mathcal{F}(A_{t-1}),$$

which proves Eq. 7. □

This invariant concerns preservation after admission; it neither establishes that one-sided projection is necessary for preservation nor guarantees a nonzero guarded update or eventual admission of an uncommitted constraint.

For a nonempty closed convex set $S$, let $P_S$ denote Euclidean projection onto $S$. For the projection-specific results below, write

$$d = \Delta z_t, \qquad J = J_t, \qquad \mathcal{C} = \{u : Ju \leq 0\}, \qquad u^{\mathrm{CCL}} = P_{\mathcal{C}}(d).$$

## A.2 Proof of Proposition 3

**Proposition 3 (Local minimum-deviation projection)** *Let*

$$\mathcal{C}_t = \{u : J_t u \leq 0\}, \qquad u_t^{\mathrm{CCL}} = P_{\mathcal{C}_t}(\Delta z_t). \tag{9}$$

*Then $u_t^{\mathrm{CCL}}$ is unique and*

$$\|u_t^{\mathrm{CCL}} - \Delta z_t\|_2 \leq \|v - \Delta z_t\|_2, \qquad \forall v \in \mathcal{C}_t. \tag{10}$$

*Moreover,*

$$J_t \Delta z_t \leq 0 \implies u_t^{\mathrm{CCL}} = \Delta z_t, \qquad \Delta z_t - u_t^{\mathrm{CCL}} \in \mathrm{range}(J_t^\top). \tag{11}$$

The set

$$\mathcal{C} = \{u : Ju \leq 0\}$$

is nonempty, closed, and convex. Its Euclidean projection therefore exists and is unique. By the defining property of projection,

$$\|u^{\mathrm{CCL}} - d\|_2 \leq \|v - d\|_2, \qquad \forall v \in \mathcal{C},$$

establishing Eq. 10.

If $Jd \leq 0$, then $d \in \mathcal{C}$, and uniqueness gives

$$u^{\mathrm{CCL}} = d.$$

For the remaining statement, the KKT conditions yield a multiplier $\lambda^\star \geq 0$ satisfying

$$u^{\mathrm{CCL}} = d - J^\top \lambda^\star.$$

Consequently,

$$d - u^{\mathrm{CCL}} = J^\top \lambda^\star \in \mathrm{range}(J^\top),$$

which proves Eq. 11. □

### A.3 Proof of Corollary 2

Let

$$\mathcal{N} = \ker J, \qquad \mathcal{L} = \mathrm{range}(J^\top).$$

Since $\mathcal{N} = \mathcal{L}^\perp$, every $u$ admits a unique decomposition $u = n + w$ with $n \in \mathcal{N}$ and $w \in \mathcal{L}$. Because $Jn = 0$,

$$Ju = Jw.$$

Hence

$$\mathcal{C} = \mathcal{N} \oplus (\mathcal{C} \cap \mathcal{L})$$

as an orthogonal direct sum. Therefore

$$P_{\mathcal{C}} d = P_{\mathcal{N}} d + P_{\mathcal{C}\cap\mathcal{L}}(P_{\mathcal{L}} d),$$

which gives

$$u^{\mathrm{CCL}} = u^{\mathrm{eq}} + q, \qquad q = P_{\mathcal{C}\cap\mathcal{L}}(P_{\mathcal{L}} d).$$

Since $\mathcal{C} \cap \mathcal{L}$ is a closed convex cone, Moreau's decomposition gives

$$\|P_{\mathcal{L}} d\|_2^2 = \|P_{\mathcal{L}} d - q\|_2^2 + \|q\|_2^2.$$

Moreover,

$$\|d - u^{\mathrm{eq}}\|_2^2 = \|P_{\mathcal{L}} d\|_2^2, \qquad \|d - u^{\mathrm{CCL}}\|_2^2 = \|P_{\mathcal{L}} d - q\|_2^2.$$

Subtracting yields Eq. 8. □

### A.4 First-Order Compatibility with Squared Radial Residuals

**Proposition 4 (First-order compatibility with squared radial residuals)** *Consider the trajectory-proxy instantiation in Section 3.2. Suppose $z_t \in \mathcal{F}(A_t)$, the references $r_k$ and targets $\tau_k$ are held fixed, and every committed position readout $q_k$, $k \in A_t$, is differentiable at $z_t$. Let $J_t$ be the exact exposed Jacobian, with the orientations in Eq. 2 held fixed during differentiation. Define*

$$g_k(z) = \frac{1}{2} c_k(z)^2 = \frac{1}{2} \|e_k(z)\|_2^2.$$

*Then every $u \in \mathcal{C}_t$ satisfies*

$$Dg_k(z_t)[u] \le 0, \qquad k \in A_t.$$

*If $c_k(z_t) > 0$, then $c_k$ is differentiable at $z_t$ and*

$$Dc_k(z_t)[u] = \frac{Dg_k(z_t)[u]}{c_k(z_t)} \le 0.$$

*Moreover, this last inequality is strict if $J^{\mathrm{pos}}_{k,a}(z_t)u < 0$ for at least one coordinate $a \in \{x, y\}$ satisfying $e_{k,a}(z_t) \neq 0$.*

**Proof.** For each $k \in A_t$, both position relations are exposed because $S^{\mathrm{pos}}_{k,a} = \{k\}$. Writing $\sigma^{\mathrm{pos}}_{k,a}(t)$ for their current-state orientations, their oriented Jacobian rows satisfy

$$J^{\mathrm{pos}}_{k,a}(z_t) = \sigma^{\mathrm{pos}}_{k,a}(t) \nabla e_{k,a}(z_t)^\top.$$

Since

$$e_{k,a}(z_t) = |e_{k,a}(z_t)| \sigma^{\mathrm{pos}}_{k,a}(t),$$

including under the declared positive-branch convention at zero, every $u \in \mathcal{C}_t$ satisfies

$$\begin{aligned} Dg_k(z_t)[u] &= \sum_{a \in \{x,y\}} e_{k,a}(z_t) \nabla e_{k,a}(z_t)^\top u \\ &= \sum_{a \in \{x,y\}} |e_{k,a}(z_t)| J^{\mathrm{pos}}_{k,a}(z_t) u \le 0. \end{aligned}$$

When $c_k(z_t) > 0$, the chain rule gives

$$Dc_k(z_t)[u] = \frac{Dg_k(z_t)[u]}{c_k(z_t)}.$$

If at least one term with positive weight is strictly negative, the corresponding sum is strictly negative. Additional exposed temporal rows only restrict the admissible set of directions. □

At $e_k(z_t) = 0$, the squared-residual statement remains valid but does not assert a nonpositive directional derivative for the nonsmooth norm $c_k$. If $\epsilon_k > 0$, such a gate is strictly interior and is instead covered by the continuity condition in Lemma 5. At a boundary gate with $c_k(z_t) = \epsilon_k > 0$, the positive-step lemma requires the strict case above. Observer-support stability and finite-step feasibility remain subject to direct evaluation by the nonlinear guard.

### A.5 Sufficient Conditions for Accepting a Positive Step Size

The commitment invariant does not require the guard to accept a positive step size, since it may select $\alpha_t^\star = 0$. The following result gives sufficient local conditions under which the finite-step guard accepts some $\alpha > 0$.

**Lemma 5 (Acceptance of a positive candidate step size)** *For each $i \in A_t$, define*

$$\phi_i(\alpha) = c_i(z_t + \alpha u_t^{\mathrm{CCL}}), \qquad m_i = \epsilon_i - \phi_i(0).$$

*Suppose $A_t$ is finite and each $\phi_i$ is continuous in a right neighborhood of zero. Assume also that observer support is locally stable along the candidate ray: for every $i \in A_t$, there exists $\delta_i^\chi > 0$ such that*

$$\chi(z_t + \alpha u_t^{\mathrm{CCL}}; r_i) = 1, \qquad \forall \alpha \in [0, \delta_i^\chi].$$

*Assume further that every committed gate is either strictly interior,*

$$m_i > 0,$$

*or lies on the boundary and has a strictly negative right derivative,*

$$m_i = 0, \qquad \phi_i'(0^+) < 0.$$

*Then there exists $\bar{\alpha} > 0$ such that*

$$z_t + \alpha u_t^{\mathrm{CCL}} \in \mathcal{F}(A_t), \qquad \forall \alpha \in [0, \bar{\alpha}].$$

*If $\Gamma \neq \varnothing$ and $\min \Gamma \leq \bar{\alpha}$, the finite-step guard accepts a positive step size. If, in addition, $u_t^{\mathrm{CCL}} \neq 0$, then $z_{t-1} \neq z_t$.*

**Proof.** If $A_t = \varnothing$, the claim is immediate. For a strictly interior gate, continuity gives a positive neighborhood on which the residual remains feasible. For a boundary gate with $\phi_i'(0^+) < 0$, the definition of the right derivative gives

$$\phi_i(\alpha) < \phi_i(0) = \epsilon_i$$

for all sufficiently small positive $\alpha$. Intersecting these residual-feasibility neighborhoods with the corresponding support-stability neighborhoods and using finiteness of $A_t$ gives a common $\bar{\alpha} > 0$. Hence

$$z_t + \alpha u_t^{\mathrm{CCL}} \in \mathcal{F}(A_t)$$

for every $\alpha \in [0, \bar{\alpha}]$. Under the grid condition, $\min \Gamma \in \Gamma \cap (0, \bar{\alpha}]$ is feasible. Therefore, by Eq. 6,

$$\alpha_t^\star \geq \min \Gamma > 0.$$

If $u_t^{\mathrm{CCL}} \neq 0$, it follows that

$$z_{t-1} - z_t = \alpha_t^\star u_t^{\mathrm{CCL}} \neq 0.$$

□

If, in addition, $c_i$ is differentiable and $\nabla c_i$ is $L_i$-Lipschitz in a neighborhood of the candidate ray, the standard smoothness bound gives

$$\phi_i(\alpha) \leq \phi_i(0) + \alpha \nabla c_i(z_t)^\top u_t^{\mathrm{CCL}} + \frac{L_i}{2} \alpha^2 \|u_t^{\mathrm{CCL}}\|_2^2.$$

This bound controls only the residual conditions; a radius for the full observer-supported guard must also lie within the support-stability intervals. This refinement is conditional. In particular, $c_k(z) = \|e_k(z)\|_2$ need not be differentiable at $e_k(z) = 0$. The finite-step guard, rather than this smooth local bound, remains the operative feasibility test.

### A.6 Additional Rank-One Local Geometry

The rank-one case gives a useful local algebraic interpretation but does not predict endpoint behavior. Suppose

$$J = \beta j^\top, \qquad \beta \neq 0, \qquad j \neq 0, \qquad a = j^\top d.$$

If $\beta$ contains both positive and negative entries, the inequalities force

$$j^\top u = 0,$$

so one-sided and exact equality protection coincide.

If all nonzero entries of $\beta$ have the same sign $s \in \{-1, +1\}$, then

$$u^{\mathrm{CCL}} = d - \frac{s[sa]_+}{\|j\|_2^2} j, \qquad u^{\mathrm{eq}} = d - \frac{a}{\|j\|_2^2} j,$$

where $[x]_+ = \max(x, 0)$. Thus, a common row orientation yields a half-space constraint, whereas opposing orientations reduce the feasible set to the equality subspace $j^\top u = 0$. This result is purely local and does not imply any ordering of rollout endpoints.

## B CCL Instantiation and Numerical Realization

### B.1 Observer Evaluation and Structured Relations

The declared observer is evaluated at the state appropriate to each part of a CCL transition. In the line-scaled realization, protection relations are linearized at the current denoising state $z_t$, and the finite-step guard evaluates committed scalar gates on candidate states

$$z_t + \alpha u_t^{\mathrm{CCL}}.$$

For a complete-action family, the guard instead evaluates the endpoint of each candidate action. Post-update certification and admission are evaluated on the realized state $z_{t-1}$; final metrics are evaluated at the endpoint specified by each protocol.

The primary matched-satisfaction CogVideoX instantiation maps eight latent temporal positions to 32 decoded frames for observer evaluation. The primary observer instantiates the position readout $q_k(z)$ in Section 3.2 as the mean-channel spatial-softmax centroid, with $\beta = 50$ on $128 \times 128$ decoded frames. This reference-free observer takes $r_k = \varnothing$ and uses the trivial support predicate $\chi(z; r_k) \equiv 1$.

Let $\mathcal{J}$ index the finite family of differentiable signed protection relations $h_j(z)$, and let $S_j \subseteq \mathcal{P}$ denote the commitment support of relation $j$. The committed set $A_t$ determines the exposed relations through

$$\mathcal{R}_t = \{j \in \mathcal{J} : S_j \subseteq A_t\}. \tag{12}$$

For the trajectory instantiation, each support contains the position constraints of the frames involved in the relation. For each coordinate $a \in \{x, y\}$, the normalized relations and their commitment supports are

$$\begin{aligned} h_{k,a}^{\mathrm{pos}}(z) &= e_{k,a}(z), & S_{k,a}^{\mathrm{pos}} &= \{k\}, \\ h_{k,a}^{\mathrm{vel}}(z) &= \frac{e_{k+1,a}(z) - e_{k,a}(z)}{\sqrt{2}}, & S_{k,a}^{\mathrm{vel}} &= \{k, k+1\}, \\ h_{k,a}^{\mathrm{acc}}(z) &= \frac{e_{k+1,a}(z) - 2e_{k,a}(z) + e_{k-1,a}(z)}{\sqrt{6}}, & S_{k,a}^{\mathrm{acc}} &= \{k-1, k, k+1\}. \end{aligned} \tag{13}$$

The 32 framewise position residuals define the certification and guard gates. The structured protection family contains

$$64 \text{ position} + 62 \text{ velocity} + 60 \text{ acceleration} = 186$$

scalar relations. Position relations use $k = 0, \ldots, 31$, velocity relations use $k = 0, \ldots, 30$, and acceleration relations use $k = 1, \ldots, 30$.

The current-state outward branch in Eq. 2 is held fixed during differentiation. At an exact zero crossing, CCL uses the deterministic positive branch; no two-sided first-order claim is made at that point. For the differentiable centroid realization used in the primary matched-satisfaction study (S1), gradients pass through the frozen decoder and the declared observer only.

**Cutie readouts.** The dwell-confirmed admission study uses a nondifferentiable Cutie (Cheng et al., 2024) observer for post-update admission and, once a commitment is active, for exact nonlinear guard evaluation. The injected pulse affects one designated admission readout only; all guard and final evaluations use clean Cutie observations. The endpoint-warp diagnostic in Appendix C.6 uses Cutie only for evaluation.

In the dwell and endpoint-warp studies, a first-frame reference mask is frozen for each seed before any arm is run and bound to object ID 1. Each 49-frame readout starts from a fresh inference core. The reference mask and object ID are supplied on the first frame only; Cutie then propagates the same object track through the remaining frames without rebinding. The core is discarded after each readout, so no tracking state is shared across readouts or arms. In the admission study, for a gate indexed by frame $k$, $\chi(z; r_k) = 1$ if and only if Cutie returns a nonempty mask for object ID 1; when supported, $m(z; r_k)$ is its centroid in normalized coordinates with the origin at the top left. An empty mask therefore fails closed. Support depends only on mask non-emptiness; no additional confidence, area, shape, or open-world semantic criterion is applied.

**Endpoint warp in the referenced-object study.** Let $\bar{z}$ denote the latent after the final ordinary denoising transition. The map $W_1(\bar{z})$ applies an integer-index vertical warp with a rounded, symmetric sinusoidal temporal schedule, zero displacement at the temporal endpoints, and a maximum displacement of one latent row. The warp uses clamped boundary indexing, without interpolation or wraparound, and $W_0(\bar{z}) = \bar{z}$. Positive $y$ denotes increasing row coordinate under the top-left image convention. Neutral uses $W_0(\bar{z})$, whereas Control uses $W_1(\bar{z})$.

### B.2 Eligibility and Post-Update Admission

The primary 32-cell CogVideoX instantiation uses the temporal prerequisite chain

$$0 \to 1 \to \cdots \to 31.$$

Constraint 0 has no prerequisite, while

$$\mathrm{Pre}(i) = \{i - 1\}, \qquad i > 0.$$

Thus, in this instantiation, $E_i(A, z) = \mathbf{1}[\mathrm{Pre}(i) \subseteq A]$ at every denoising transition; no denoising-step-dependent eligibility mask is applied. Here $i$ indexes a trajectory constraint rather than a denoising step. Admission scans this declared order and adds each qualifying constraint to the temporary committed set before evaluating later constraints. A newly admitted prerequisite can therefore enable downstream constraints within the same realized state.

This ordered immediate-write policy is an instantiation of the general state-dependent admission rule; it is not a claim that CogVideoX denoises frames causally.

For each seed in the primary matched-satisfaction study (S1), the parabolic replication, and the auxiliary matched-satisfaction target-family studies, let $z_{\mathrm{init}}$ denote the seed-specific initial latent state shared across paired arms. Targets are deterministically blended toward the observer readout at this state so that

$$\max_k c_k(z_{\mathrm{init}}) \le 0.8\,\epsilon.$$

This calibration is specific to the matched-satisfaction experiments and is not used to construct the targets in the preregistered staggered-admission study.

The primary matched-satisfaction CogVideoX schedule begins with a six-transition shared hold at $z_{\mathrm{init}}$, implemented as

$$\Delta z_t^{\mathrm{init}} = g(z_{\mathrm{init}} - z_t), \qquad g = 0.3.$$

Because the window begins at $z_t = z_{\mathrm{init}}$, this recurrence leaves the state unchanged throughout the window. During the remaining 24 transitions, all paired arms scale the denoiser proposal by the shared factor $\gamma = 0.5$. The aligned-free reference is the unconstrained rollout under this calibrated sampling schedule rather than the default CogVideoX trajectory.

### B.3 One-Sided Projection and Numerical Validation

The zero-slack CCL projection used in the primary matched-satisfaction execution is solved through the nonnegative row-space dual

$$\lambda^\star = \arg\min_{\lambda \geq 0} \frac{1}{2}\lambda^\top G_t \lambda - (J_t \Delta z_t)^\top \lambda, \qquad G_t = J_t J_t^\top, \tag{14}$$

with primal recovery

$$u_t^{\mathrm{CCL}} = \Delta z_t - J_t^\top \lambda^\star.$$

For the primary matched-satisfaction CogVideoX execution, the Jacobian provider materializes and caches in float32 the Jacobian of the full structured-relation family with respect to the current latent state using reverse-mode automatic differentiation. The current commitment state selects the rows indexed by $\mathcal{R}_t$ to form $J_t$. The selected Jacobian and proposal are upcast to float64 before constructing $J_t \Delta z_t$ and $G_t = J_t J_t^\top$.

The solver attempts an active-set method, cyclic coordinate descent, and SciPy L-BFGS-B in sequence, accepting a candidate only after the same KKT check. If none passes, the zero-slack dense solver uses the equivalent SciPy NNLS formulation as its final fallback. These numerical paths solve the same projection problem rather than defining different control objectives.

**Factorized implementation.** The Full CCL arm in the held-out constraint-handler benchmark and the exact-factorized entry in the single-seed Jacobian-realization comparison use the factorized numerical realization below. Let $e(z_t) \in \mathbb{R}^{64}$ collect the position-error coordinates and define

$$J_{e,t} = \frac{\partial e(z_t)}{\partial z_t}.$$

The exposed structured Jacobian admits the exact factorization

$$J_t = B_t J_{e,t}, \tag{15}$$

where $B_t$ encodes the current outward orientations, normalized temporal stencils, and lifecycle-conditioned row exposure. In the screened runtime comparison, the backend evaluates $J_t \Delta z_t$ without materializing $J_{e,t}$ until the first directional-screen failure, leaving a proposal that passes the tolerance-aware screen unchanged. At the first failure and on every subsequent controlled transition, it forms $B_t$ and $J_{e,t}$ and solves in factorized form; the dense product $B_t J_{e,t}$ is materialized only if the final zero-slack NNLS fallback is invoked. Within each protocol, the factorized realization preserves the observer evaluated on VAE-decoded frames, the corresponding target-construction rule, the 186-relation family, the commitment lifecycle, the nonlinear guard, and the one-sided projection objective. It is nevertheless numerically distinct from the float32 direct-row realization used in the primary matched-satisfaction study, which differentiates the structured rows independently.

The solver interface uses a base tolerance of $10^{-4}$, augmented in the dense certificate by a solver-input-dtype numerical allowance. Candidate dual solutions are checked for primal feasibility, dual feasibility, active-set residuals, and complementarity; stationarity of the recovered primal update holds by construction. One-sided feasibility is then re-evaluated in the scheduler execution dtype, after casting when necessary and before the finite-step guard is evaluated. This is an execution-dtype feasibility check, not a check performed after the denoising transition has been executed.

If no candidate solution passes these numerical checks, the run aborts; solver failure is never converted into a guarded zero-step transition. By contrast, $\alpha_t^\star = 0$ is a legitimate fail-closed guard outcome after a valid projected update has been obtained.

### B.4 Lifecycle-Matched CCL–ExactSym

CCL–ExactSym changes only the local protection operator. Given the exposed Jacobian $J_t$, it replaces

$$J_t u \leq 0$$

with

$$J_t u = 0$$

and computes

$$u_t^{\mathrm{ExactSym}} = P_{\ker J_t}(\Delta z_t) = \Delta z_t - J_t^\top (J_t J_t^\top)^\dagger J_t \Delta z_t. \tag{16}$$

For the primary matched-satisfaction CogVideoX execution, $J_t$ is selected from the cached float32 full-family Jacobian and upcast to float64 before constructing and symmetrizing

$$G_t = J_t J_t^\top.$$

The eigendecomposition uses `torch.linalg.eigh`. Numerical rank uses the larger of a singular-value-relative cutoff of $10^{-6}$, squared in Gram-eigenvalue space, and the implementation's float64 roundoff floor. No damping is used.

Numerical rank, retained eigenmodes, and equality-projection residuals are validated. Failure of these checks aborts the formal run. CCL and CCL–ExactSym otherwise share the lifecycle, admission rule, structured relations, support-conditioned exposure, and finite-step guard, but evolve independently after their executed updates diverge.

### B.5 Scalar Full-Gate Bisection

Scalar full-gate bisection is a direction-only comparator governed by a single rollout-level activation flag. Before the full gate

$$\max_k c_k(z_t) \le \epsilon$$

is first satisfied, the proposal passes through unchanged. At that transition and thereafter, the arm evaluates

$$z_t + \alpha \Delta z_t, \qquad \alpha \in [0, 1].$$

It accepts $\alpha = 1$ when feasible. If $\alpha = 1$ is infeasible but $\alpha = 0$ is feasible, it performs 16 bracketed bisection iterations and returns the feasible lower endpoint. If $\alpha = 0$ is infeasible, it returns $0$. This is a finite-resolution bracketed procedure; because monotonicity of nonlinear feasibility along the ray is not established, it is not a verified global maximization of $\alpha$.

### B.6 Held-Out Constraint-Handler Realizations

The Full CCL arm in the held-out constraint-handler benchmark uses the factorized rowwise realization in Eq. 15; it is numerically distinct from the float32 direct-row realization used in the primary matched-satisfaction study. DirectGate-Jacobian retains irreversible FirstHit admission and the fixed finite-step guard, but replaces the 186 structured protection rows with 32 unnormalized radial rows,

$$g_k(z) = \tfrac{1}{2}\|e_k(z)\|_2^2, \qquad J_{g_k}(z) = e_k(z)^\top \frac{\partial e_k(z)}{\partial z}.$$

The nonlinear gate remains $\|e_k(z)\|_2 \le \epsilon$. An exactly zero radial row is omitted from projection but retained in the nonlinear guard.

Guard-only Continuous applies no projection. It first verifies $\alpha = 0$ and then tests $j/32$, $j = 1, \ldots, 32$, in ascending order until the first infeasible sample. If all sampled points are feasible, it returns $1$; otherwise, it refines the first feasible–infeasible bracket by at most 32 bisection iterations, stopping when the bracket width is at most $10^{-6}$, and returns the verified feasible lower endpoint. It does not inspect later feasible intervals or certify a global maximizer along the ray.

TR-SQP starts from the feasible zero increment and minimizes

$$\tfrac{1}{2}\|u - \Delta z_t\|_2^2$$

under the currently admitted radial gates. Its initial trust radius is $\max\{2\|\Delta z_t\|_2, 10^{-5}\}$, with configured caps of eight iterations and 24 nonlinear evaluations per transition. It returns the best verified feasible incumbent, or the zero increment if no better incumbent is found.

The Static-All Restoration handler in the held-out benchmark is distinct from the exposure-only Static-All diagnostic in the staggered study. It exposes all 186 structured rows. When the violation gradient is nonzero, it adds to the raw proposal a unit-gain restoration direction obtained from the

normalized negative gradient of the summed squared normalized positive violations, with norm $\max\{\|\Delta z_t\|_2, 10^{-6}\}$. It then projects the resulting proposal against all 186 rows and tests

$$\left\{1, \tfrac{1}{2}, \tfrac{1}{4}, \tfrac{1}{8}, \tfrac{1}{16}\right\}.$$

When the current state contains violations, it accepts the first candidate that preserves currently satisfied gates and lexicographically reduces the maximum and then the sum of normalized violations. When the current state is feasible, acceptance instead requires all-gate feasibility. If no candidate qualifies, the handler executes the zero increment.

The Reversible-Dwell-Hysteresis comparator has four states: inactive, candidate, active, and released. A gate becomes a candidate when its residual is at most $0.9\epsilon$ and either it is the root gate or its predecessor is active at the start of the current admission update. It becomes active after two consecutive qualifying admission checks. Only active gates are protected and count as prerequisites for downstream gates. Active gates are guarded at $1.2\epsilon$ and released after two consecutive residuals above $\epsilon$. When one or more gates meet the release criterion, the earliest such gate and every downstream gate are reset to the released state. In the selected configuration, a released gate returns to the inactive state when its two-update cooldown counter expires.

For a common auxiliary diagnostic, the realized states from all six arms are additionally scored under the same method-independent FirstHit rule; this scoring does not affect any arm's control path.

### B.7 Transition Realizations and Experimental Constants

For the line-scaled realization, Algorithm 1 records the ordering essential to CCL: existing commitments determine exposure and protection, the projected update is checked at finite scale, execution realizes the next state, admission evidence is updated, and only then are newly qualifying constraints admitted.

For Algorithm 1, let

$$b_i(z) = \mathbf{1}[\chi(z; r_i) = 1 \text{ and } c_i(z) \le \epsilon_i]$$

indicate supported gate satisfaction, and let $E_i(A, z) \in \{0, 1\}$ indicate eligibility for admission given the committed set $A$ and state $z$.

Here $K \ge 1$ is an integer dwell length. For each uncommitted constraint $i$, $\ell_{i,t}$ records the number of consecutive realized post-update states at which its gate is supported and satisfied, capped at $K$. The vector $\ell_t$ collects these counts.

**Algorithm 1** One line-scaled CCL denoising transition

**Require:** $z_t, \Delta z_t, A_t, \ell_t, K, \{b_i, E_i\}_{i\in\mathcal{P}}, \{h_j, S_j\}_{j\in\mathcal{J}}, \Gamma$; $z_t \in \mathcal{F}(A_t)$

$\mathcal{R}_t \leftarrow \{j \in \mathcal{J} : S_j \subseteq A_t\}$
**if** $\mathcal{R}_t = \emptyset$ **then**
  $u_t^{\mathrm{CCL}} \leftarrow \Delta z_t$
**else**
  form the current outward relations and exposed Jacobian $J_t = \mathrm{stack}_{j\in\mathcal{R}_t} J_j(z_t)$
  solve Eq. 3 for $u_t^{\mathrm{CCL}}$
**end if**
$\alpha_t^\star \leftarrow \max\Big(\Big\{\alpha \in \Gamma : b_i\Big(z_t + \alpha u_t^{\mathrm{CCL}}\Big) = 1 \ \forall i \in A_t\Big\} \cup \{0\}\Big)$
$z_{t-1} \leftarrow z_t + \alpha_t^\star u_t^{\mathrm{CCL}}$
**for** $i \in \mathcal{P} \setminus A_t$ **do**
  $\ell_{i,t-1} \leftarrow b_i(z_{t-1}) \min\{K, \ell_{i,t} + 1\}$
**end for**
$\widetilde{A} \leftarrow A_t$
**for** $i \in \mathcal{P} \setminus A_t$ in declared topological order **do**
  **if** $E_i(\widetilde{A}, z_{t-1}) = 1$ and $\ell_{i,t-1} = K$ **then**
    $\widetilde{A} \leftarrow \widetilde{A} \cup \{i\}$
  **end if**
**end for**
$A_{t-1} \leftarrow \widetilde{A}$
$\ell_{t-1} \leftarrow (\ell_{i,t-1})_{i\in\mathcal{P}\setminus A_{t-1}}$
**return** $z_{t-1}, A_{t-1}, \ell_{t-1}$

Table 5: **Primary matched-satisfaction CogVideoX-2b configuration.**

| Quantity | Value / convention |
|---|---|
| Backbone | frozen `THUDM/CogVideoX-2b`; revision `1137dacfc2c9c012bed6a0793f4ecf2ca8e7ba01` |
| Sampling schedule | 30 scheduled transitions; denoiser guidance scale 6.0 |
| Proxy gate | $\epsilon = 0.05$ |
| Observer | spatial-softmax centroid; $\beta = 50$; $128 \times 128$ |
| Temporal resolution | 8 latent frames $\rightarrow$ 32 observer frames |
| Structured relations | 64 position, 62 velocity, 60 acceleration; 186 total |
| Matched-target calibration | $\max_k c_k(z_{\text{init}}) \leq 0.8\epsilon$ |
| Initial hold | first 6 transitions at $z_{\text{init}}$; $g = 0.3$ |
| Denoiser-proposal scaling | $\gamma = 0.5$ over the remaining 24 transitions, shared by matched-satisfaction arms |
| CCL/CCL–ExactSym guard schedule | $\{1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}\}$, then 0 |
| CCL/CCL–ExactSym eligibility and admission | prerequisite-only, with no denoising-step mask; post-update with $K = 1$, irreversible, ordered immediate-write |
| CCL/CCL–ExactSym row exposure | $S_j \subseteq A_t$ |
| Primary matched-satisfaction Jacobian path | full-family float32 reverse-mode AD; $\mathcal{R}_t$-selection; float64 row-space solve |
| CCL base tolerance | $10^{-4}$, plus dtype-dependent validation allowance |
| ExactSym | cached float32 Jacobian copied to float64; float64 Gram eigendecomposition; rank rtol $10^{-6}$; float64 roundoff floor; zero damping |
| Scalar | pass-through until first full-gate satisfaction; thereafter, 16-step bracketed ray bisection returning the feasible lower endpoint |
| Final reporting gate | $\max_k c_k(z_{\text{final}}) \leq \epsilon + 10^{-3}$ |

# C Evaluation Protocols and Statistical Definitions

## C.1 Method-Blind Staggered Admission

For each formal seed, $q_{\text{init}} \in \mathbb{R}^{32\times 2}$ denotes the observer readout at the initial aligned-free state, and $q_{\text{post}}[a] \in \mathbb{R}^{32\times 2}$ denotes the cached aligned-free post-update readout at index $a$. Before any controlled arm is run, these readouts are used once to construct a seed-specific target, which is then held fixed across arms and denoising steps. The 32 certification cells are partitioned into four contiguous eight-cell blocks $B_0, \ldots, B_3$. For every strictly increasing triple $6 \leq a_1 < a_2 < a_3 \leq 28$, the construction sets

$$q_{\text{target},B_0} = q_{\text{init},B_0}, \qquad q_{\text{target},B_j} = q_{\text{post}}[a_j, B_j], \quad j = 1, 2, 3.$$

The prespecified target-construction screen evaluates all 1,771 triples using only these aligned-free readouts. A triple is retained only if $B_0$ is initially satisfied; every delayed block contains at least one cell with initial residual greater than $\epsilon + 0.005$; prerequisite-chain replay completes the delayed blocks no earlier than step 6, with adjacent completion gaps of at least two steps; at least eight cells subsequently exceed $\epsilon + 0.005$ after their first replayed hit; and at least one transition remains after final completion. Passing triples are ranked first by decreasing minimum temporal slack, then by decreasing minimum number of initially outside cells across the delayed blocks, and then by decreasing number of regression-opportunity cells. Remaining ties are broken in favor of the lexicographically earliest anchor tuple.

Formal seeds are 7520–7551, with no replacement, and all 32 pass the prespecified target-construction screen. The selected anchors are offline target-construction metadata, not runtime eligibility thresholds or prescribed admission times. Runtime eligibility remains prerequisite-only, with no denoising-step mask; realized admission is determined by post-update gate satisfaction and the declared prerequisite chain.

The four factorial arms use the same targets and post-update certification timing. Certify-only executes the raw proposal without protection; Projection-only applies one-sided projection without

the finite-step guard; Guard-only applies the guard to the raw proposal; and CCL applies both projection and guarding.

The preregistered staggered-admission protocol also includes a secondary Static-All row-exposure diagnostic. It retains the commitment lifecycle, prerequisite-only eligibility, post-update admission, one-sided projection, and finite-step guard, but exposes the full structured-relation family from the first controlled transition,

$$\mathcal{R}_t^{\mathrm{Static}} = \mathcal{J}.$$

This diagnostic changes only relation exposure: it does not initialize all constraints as committed, and the guard evaluates only admitted scalar gates. It is distinct from Static-All Restoration in the held-out benchmark, and its results are interpreted descriptively.

This protocol is therefore a mechanism-isolation test of lifecycle behavior, not a test of acquiring trajectories specified independently of the aligned-free rollout.

### C.2 Exploratory Held-Out Constraint-Handler Benchmark

The exploratory benchmark evaluates 64 distinct held-out seeds 731100–731163, with no replacement. Each seed is evaluated under six handlers: Full CCL, TR-SQP, DirectGate-Jacobian, Static-All Restoration, Guard-only Continuous, and Reversible-Dwell-Hysteresis. The handlers share the frozen model revision, prompt, initial-latent seed, 30-transition schedule, observer, 32 certification gates, and a seed-specific method-blind staggered target constructed once from the cached aligned-free readouts. The target is fixed across handlers and denoising steps, whereas the six rollouts evolve independently. Handler configurations and numerical budgets are selected on disjoint development seeds and fixed before held-out execution. All handlers execute the full rollout; zero updates permitted by a handler are retained as algorithmic outcomes.

Full CCL uses the FirstHit lifecycle, the lifecycle-conditioned family of 186 structured position, velocity, and acceleration relations, one-sided projection, and the finite-step guard. DirectGate-Jacobian retains the lifecycle and guard but replaces the structured relations with the direct squared-error rows $e_k^\top J_{e_k}$ for the active scalar gates. TR-SQP enforces the active scalar gates under its frozen trust-region budget. Static-All Restoration exposes all 186 relations from the first transition and includes a violation-restoration update. Guard-only Continuous searches the unprojected proposal ray from zero and refines the first sampled feasible–infeasible bracket. Reversible-Dwell-Hysteresis instead uses a frozen admission–release policy with dwell, hysteresis, and cooldown; it is distinct from the irreversible dwell rule evaluated later in this section.

Let $r \in \{1, \dots, 30\}$ denote the ordinal of the realized post-update state within the rollout.

For seed $s$ and handler $m$, terminal-window success is

$$Y_s^{(m)} = \mathbf{1}\left[c_k\left(z_{s,r}^{(m)}\right) \le \epsilon \quad \forall k \in \{0, \dots, 31\}, \quad \forall r \in \{28, 29, 30\}\right].$$

The 186 structured relations define protection geometry but are not scored as task outcomes. The statistical unit is the seed.

The normalized intervention is

$$I_s^{(m)} = \frac{\sum_t \left\| v_{s,t}^{\mathrm{exec},(m)} - d_{s,t}^{\mathrm{raw},(m)} \right\|_2^2}{\sum_t \left\| d_{s,t}^{\mathrm{raw},(m)} \right\|_2^2 + 10^{-12}},$$

where the raw proposal is computed at that handler's own current state. Success-rate differences are defined as Full CCL minus the comparator and are reported with $95\%$ Tango paired efficient-score confidence intervals (Tango, 1998). Exact two-sided McNemar tests use the paired terminal outcomes, with Holm adjustment across the five comparator tests.

Intervention differences are estimated only among seeds on which both handlers succeed, using 20,000 paired seed-level percentile bootstrap draws. Negative differences indicate lower intervention under Full CCL. Because common success is a post-treatment subset, these conditional summaries

are descriptive. The benchmark reports exploratory effect estimates under frozen configurations and budgets rather than a confirmatory superiority test.

### C.3 Matched Satisfaction

The primary matched-satisfaction study (S1) and its target-family replications use deterministic calibration to define an initially feasible preservation setting. Each seed-specific target is moved toward the initial observer readout until

$$\max_k c_k(z_{\text{init}}) \le 0.8\,\epsilon.$$

For the primary family, $\widetilde{q}$ is obtained by separate linear least-squares fits to the $x$- and $y$-coordinates of $q_{\text{init}}$, evaluated at normalized frame times. For the parabolic family, the $x$-coordinate is fit linearly and the $y$-coordinate quadratically at the same times. After $\widetilde{q}$ is clipped coordinatewise to $[0.02, 0.98]$, the target is

$$q_{\text{target}} = q_{\text{init}} + \lambda\big(\widetilde{q} - q_{\text{init}}\big),$$

where $\lambda \in (0, 1]$ is the largest scalar satisfying the initial calibration condition above and is then held fixed across arms.

For methods $m_1$ and $m_2$, define

$$\mathcal{S}_{\text{match}}(m_1, m_2) = \left\{ s : \max_k c_k(z^{m,s}_{\text{final}}) \le \epsilon + 10^{-3}, \quad \forall m \in \{m_1, m_2\} \right\}. \tag{17}$$

The $10^{-3}$ tolerance is used only for final reporting and matched-set construction; certification and guard decisions use $\epsilon$.

Paired methods share prespecified exogenous inputs and target metadata but evolve independently. The aligned-free counterpart is the final output of the unconstrained rollout under the same calibrated sampling configuration. The resulting estimand is collateral deviation conditional on matched final proxy satisfaction.

### C.4 Deviation Metrics and Statistical Reporting

Let $D_x(m, s)$ denote the distance between method $m$'s final output for seed $s$ and its aligned-free counterpart. Pixel deviation is RMSE between the final decoded video tensors. Latent deviation is the unnormalized $\ell_2$ distance between the final latent states. LPIPS (Zhang et al., 2018) is computed with the AlexNet backbone, and the DINO (Caron et al., 2021) feature distance uses `facebook/dino-vits16`. Both are evaluated between corresponding decoded frames and averaged over frames. All four are endpoint metrics; none summarizes the intermediate denoising path.

For comparator $b$ and metric $x$, relative reduction is computed on

$$\mathcal{S}^x_{\text{rel}}(\text{CCL}, b) = \big\{ s \in \mathcal{S}_{\text{match}}(\text{CCL}, b) : D_x(b, s) > 10^{-12} \big\}$$

as

$$\text{Red}_x(\text{CCL}, b) = 100 \operatorname*{median}_{s \in \mathcal{S}^x_{\text{rel}}} \frac{D_x(b, s) - D_x(\text{CCL}, s)}{D_x(b, s)}. \tag{18}$$

This is the median of the per-seed reductions, not a reduction computed from the displayed group medians.

The paired mean difference is computed over the full matched set:

$$\Delta_x(s) = D_x(b, s) - D_x(\text{CCL}, s),$$

so positive values favor CCL. Intervals for the primary matched-satisfaction study use 10,000 paired seed-level bootstrap draws. Each draw resamples seed indices with replacement, preserves method

pairing, and recomputes the mean paired difference; the reported interval is given by the 2.5th and 97.5th percentiles. The seed is the resampling unit.

A supplementary analysis, reported in Table 10, uses 100,000 paired seed-level bootstrap draws for the median relative reductions and 1,000,000 Monte Carlo sign-flip draws for the mean differences. The four metrics form one family for CCL versus Scalar and a separate exploratory family for CCL versus CCL–ExactSym; Holm correction is applied within each family. Because these procedures target different estimands, neither substitutes for the other.

For the Static-All row-exposure diagnostic, final-gate outcomes are paired with CCL on the same 32 formal seeds. The one-sided exact McNemar $p$-value is reported descriptively because no separate confirmatory contrast was prespecified. Endpoint distances are evaluated only on the common-success subset and are likewise treated as descriptive post-treatment comparisons.

### C.5 Dwell-Confirmed Admission Protocol

This separate stress test evaluates whether an isolated supported crossing can be written irreversibly into the commitment state and whether consecutive-hit dwell prevents that failure. The earlier proxy experiments use the $K = 1$ admission rule.

For each candidate seed, cohort selection uses seven clean Cutie readouts spanning the entry state and the six-transition suffix, evaluated at decoded frame 5. The target is the mean clean centroid over states 3–5, and $\epsilon$ is the midpoint between the largest residual over states 3–5 and the smallest residual over states 0–2. Eligibility requires supported readouts at all seven states, the clean gate to be violated at states 0–2 and satisfied at states 3–5, and a clean residual-to-$\epsilon$ ratio at state 1 in $(1, 1.5]$. No pulse is constructed or evaluated during cohort selection.

The cohort comprises the first 64 clean-only eligible tasks encountered in the frozen seed order from 26000 to 26999. Each task is evaluated in six branches,

$$K \in \{1, 2, 3\} \times \{\text{clean}, \text{pulse}\},$$

over the same six-transition suffix. The reference mask and object identity are fixed, automatic rebinding is disabled, and each readout uses a fresh Cutie inference core as specified in Appendix B.

The pulse is applied only to the post-update admission readout at the first transition of the formal suffix. It uses base brightness 24, checker amplitude 8, and phase 0; its nonzero support is disjoint from the clean Cutie support dilated by five pixels. It leaves the latent state, proposal, scheduler, random-number state, reference mask, and guard readout unchanged. All guard evaluations and final evaluations use the clean Cutie readout.

A false admission occurs when the pulse branch creates an irreversible commitment at the pulse transition while the paired clean readout on the same realized state fails the gate. Clean acquisition requires runtime admission of the intended tracked object, final clean-gate satisfaction, and no unresolved guard failure.

Accordingly, this experiment is a robustness test of the admission readout and lies outside the fixed state-to-gate contract of Theorem 1.

Let

$$p_K = P(\text{FA} \mid K).$$

For $k \in \{2, 3\}$, define the paired false-admission risk reduction

$$B_k = p_1 - p_k.$$

For each contrast, inference uses the $97.5\%$ lower confidence bound from Tango's paired efficient-score interval, corresponding to a Bonferroni-adjusted one-sided level of $0.025$. We select the smallest $K$ whose lower confidence bound exceeds zero. Clean acquisition is reported using paired risk differences and nominal two-sided $95\%$ intervals; no noninferiority margin was prespecified.

### C.6 Referenced-Object Evaluation Protocol

The referenced-object study uses one frozen red-balloon task and eight held-out seed pairs,

18103, 18105, 18107, 18110, 18113, 18115, 18119, 18123,

with no replacement.

Within each pair, Neutral decodes the unmodified endpoint $\bar{z}$, whereas Control decodes $W_1(\bar{z})$, with the prespecified positive-$y$ warp applied immediately after the final ordinary transition. Neither arm uses endpoint interpolation or recovery search.

For evaluation only, a pair-specific first-frame Cutie mask for the target red balloon is frozen before either arm is run and shared across the two arms. The mask is bound to internal object ID 1, and automatic rebinding is disabled. Both arms are evaluated over all 49 frames. Cutie is used only for evaluation and does not affect endpoint construction, runtime guarding, or admission.

Using zero-based decoded-frame indices, let

$$\mathcal{F}_{\mathrm{pre}} = \{0, \ldots, 10\}, \qquad \mathcal{F}_{\mathrm{act}} = \{11, \ldots, 37\}.$$

For pair $s$, let $y^{\mathrm{C}}_{s,f}$ and $y^{\mathrm{N}}_{s,f}$ denote the reference-bound Cutie $y$-centroids after conversion to decoded-pixel coordinates, and define

$$r_{s,f} = y^{\mathrm{C}}_{s,f} - y^{\mathrm{N}}_{s,f}, \qquad b_s = \mathrm{median}_{f \in \mathcal{F}_{\mathrm{pre}}}\, r_{s,f},$$

$$\delta_s = \frac{1}{27} \sum_{f \in \mathcal{F}_{\mathrm{act}}} \big(r_{s,f} - b_s\big).$$

The directional endpoint is the mean of $\delta_s$ over the eight pairs. Positive values indicate motion in the prescribed positive-$y$ direction, and a pair is directionally positive when $\delta_s > 0$. The seed pair is the statistical unit; frames are repeated measurements within a pair. A separate non-gating, method-blind identity review records whether the tracked instance is the target balloon; it does not enter pair selection or the directional endpoint.

This protocol tests whether the measured image-plane motion remains tied to the referenced object under the endpoint warp; it does not compare Full CCL with another constraint handler.

# D Additional Lifecycle and Preservation Results

This appendix expands the lifecycle and preservation results summarized in Sections 4.1–4.2. We first examine persistence and the roles of projection and guarding, then compare alternative constraint handlers and report matched-satisfaction replications and an admission-policy stress test.

## D.1 Staggered Admission

We use the staggered-admission study to test whether recording post-update satisfaction is sufficient for persistence. Each of the 32 seeds yields a passing triple under the prespecified target-construction screen. Table 6 summarizes the lifecycle outcomes.

Table 6: **Staggered-admission lifecycle outcomes on 32 paired seeds.** Later violations are counted after first certification for Certify-only and after admission for Static-All and CCL. Static-All is the prespecified secondary row-exposure diagnostic; "–" indicates that no guard is applied.

| Outcome | Certify-only | Static-All | CCL |
|---|---|---|---|
| Final all-gate criterion | 0/32 | 17/32 | 30/32 |
| Later violating cell–steps | 7,247 | 0 | 0 |
| Seed–constraint pairs with a later violation | 927 | 0 | 0 |
| Zero guarded steps | – | 0 | 0 |

**Paired completion.** For the preregistered primary comparison, 30 seeds succeed only under CCL, none succeeds only under Certify-only, and two fail under both methods (one-sided exact McNemar

$p = 9.31 \times 10^{-10}$). The two CCL failures reflect incomplete acquisition rather than the loss of an admitted commitment.

Thus, recording a satisfaction event without carrying it forward as a protection obligation is insufficient for persistence in this cohort.

### D.2 Static-All Row Exposure

The prespecified Static-All diagnostic isolates the effect of exposing the full relation family from the first controlled transition while retaining the same commitment lifecycle and guard. Its outcomes are reported in Table 6 alongside the primary comparison.

Against Static-All, 14 seeds succeed only under CCL and one succeeds only under Static-All; the remaining pairs comprise 16 joint successes and one joint failure. The paired final-completion difference is $40.625$ percentage points in favor of CCL. The descriptive one-sided exact McNemar $p$-value is $4.88 \times 10^{-4}$; no separate confirmatory Static-All contrast was prespecified.

**Accepted step sizes.** Across the 960 controlled transitions, the accepted-$\alpha$ counts at $\left(1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, 0\right)$ are $(931, 23, 5, 1, 0, 0)$ for Static-All and $(927, 27, 5, 1, 0, 0)$ for CCL. The corresponding mean accepted step sizes are $0.9832$ and $0.9811$. Neither method takes a zero step, so the lower completion rate under Static-All does not coincide with more frequent backtracking or zero-step stalling.

**Common-success deviation.** Table 7 reports final-output deviation on the 16 seeds completed by both protected methods.

Table 7: **Common-success final-output deviation.** Mean differences are Static-All minus CCL on the 16 common-success seeds; intervals are paired seed-level percentile-bootstrap $95\%$ CIs.

| Metric | Mean difference | 95% CI |
|---|---|---|
| Pixel | 0.000570 | $[-0.002264, 0.003609]$ |
| Latent | 0.3854 | $[-3.6158, 4.5552]$ |

Both intervals include zero. Because joint completion is a post-treatment condition, these estimates describe only the observed common-success subset, not an unconditional policy effect.

Neither protected method loses an admitted commitment. Static-All attains lower completion in this cohort without additional backtracking or stalling; this secondary diagnostic does not identify an optimal relation-exposure schedule.

### D.3 Projection–Guard Factorial

The projection–guard factorial separates the roles of the two components. It reuses the 32 seeds and targets from the staggered study and is therefore a component analysis rather than an independent replication.

The first prespecified fixed-sequence contrast compares CCL with Projection-only on seed-level post-admission regression. Adding the guard reduces regression from 25/32 seeds to 0/32 (paired risk difference, Projection-only minus CCL, $0.78125$; exact two-sided McNemar $p = 5.96 \times 10^{-8}$). The second contrast compares CCL with Guard-only on final all-gate completion. Adding projection increases completion from 0/32 to 30/32 (paired risk difference, CCL minus Guard-only, $0.9375$; exact two-sided McNemar $p = 1.86 \times 10^{-9}$). Within this shared cohort, the contrasts support complementary roles: projection supports acquisition, whereas the guard prevents subsequent regression.

### D.4 Endpoint Comparison of Protection Geometries

A separate held-out component study compares five protection realizations on 32 paired seeds (7600–7631). All arms use the same observer, seed-specific targets, lifecycle, nonlinear guard, rollout

Table 8: **Endpoint comparison of protection geometries on 32 paired held-out seeds.** The primary comparison is Structured versus Gate-radial one-sided protection. The table reports the number of trajectories meeting the prespecified consistent-completion endpoint.

| Protection realization | Consistent completion |
|---|---|
| Guard-only | 0/32 |
| Gate-radial one-sided | 21/32 |
| Position one-sided | 22/32 |
| Structured one-sided | 30/32 |
| Structured ExactSym | 3/32 |

schedule, and prespecified consistent-completion endpoint. Gate-radial one-sided protection uses one radial row for each scalar gate. Position one-sided protection instead uses 64 coordinatewise position rows, while Structured one-sided protection augments these rows with 62 velocity and 60 acceleration relations. Structured ExactSym retains the same 186-relation family but replaces the one-sided inequalities with equality protection. Guard-only applies the nonlinear guard without a projection step. The arms are paired by seed but evolve independently after their executed updates diverge; the resulting contrasts are therefore endpoint comparisons rather than same-state operator comparisons.

Consistent completion requires the final committed set to contain all 32 lifecycle cells and the final denoising state $z_{\mathrm{final}}$ to satisfy

$$\max_k c_k(z_{\mathrm{final}}) \le \epsilon + 10^{-9}.$$

The seed is the statistical unit.

The prespecified primary comparison is Structured versus Gate-radial one-sided protection. The Structured-versus-Position comparison tests the incremental contribution of the velocity and acceleration relations and is reached only if the primary comparison meets its prespecified criterion. Comparisons with Structured ExactSym and Guard-only are secondary.

Structured one-sided protection completes 30/32 trajectories, compared with 21/32 for gate-radial protection, yielding a paired risk difference of 0.28125 (95% CI $[0.09375,\ 0.46875]$; exact two-sided McNemar $p = 0.0117$). The primary comparison meets its prespecified criterion, allowing the Structured-versus-Position comparison to proceed. Position-only protection completes 22/32 trajectories, giving a paired completion advantage of 0.25000 for Structured protection (95% CI $[0.09375,\ 0.40625]$; exact two-sided McNemar $p = 0.0215$). Both intervals use 10,000 paired seed-level percentile-bootstrap resamples (RNG seed 20260820). Together, these comparisons support a completion benefit from the structured geometry and an incremental contribution of its temporal relations under the shared lifecycle and nonlinear guard.

The secondary comparison with Structured ExactSym also favors one-sided protection. Structured one-sided protection completes 30/32 trajectories, compared with 3/32 under equality protection, with 28 discordant seeds favoring one-sided protection and one favoring equality protection. The difference remains significant after Holm correction within the prespecified family of secondary binary contrasts ($p_{\mathrm{Holm}} = 1.12 \times 10^{-7}$). Guard-only completes none of the 32 trajectories, a descriptive result consistent with the acquisition role of projection observed in the projection–guard factorial. Together, these results provide protocol-specific evidence for one-sided rather than equality protection under the same structured relation family.

### D.5 Held-Out Constraint-Handler Benchmark

We next compare CCL with alternative constraint handlers under prespecified, handler-specific numerical budgets. All handlers are evaluated by the same terminal-window success criterion: all 32 gates must hold at terminal states 28–30. Table 9 reports this outcome on 64 paired seeds together with paired comparisons against Full CCL.

Table 9: **Held-out constraint-handler benchmark on 64 paired seeds.** Success requires all 32 gates to hold at terminal states 28–30. Paired risk differences are Full CCL minus comparator, so positive values favor Full CCL; pp denotes percentage points. Tango 95% confidence intervals are nominal and unadjusted. Exact two-sided paired McNemar $p$-values are Holm-adjusted across the five comparisons.

| Method | Terminal-window success | Difference (pp) | Nominal Tango 95% CI (pp) | Holm-adjusted $p$ |
|---|---|---|---|---|
| Full CCL | 60/64 | – | – | – |
| Static-All Restoration | 46/64 | $+21.875$ | $[+9.054, +34.961]$ | $0.00258$ |
| DirectGate-Jacobian | 44/64 | $+25.000$ | $[+11.734, +38.283]$ | $0.00171$ |
| TR-SQP | 2/64 | $+90.625$ | $[+79.831, +95.632]$ | $2.08 \times 10^{-17}$ |
| Guard-only Continuous | 0/64 | $+93.750$ | $[+82.779, +97.543]$ | $8.67 \times 10^{-18}$ |
| Reversible-Dwell-Hysteresis | 0/64 | $+93.750$ | $[+82.779, +97.543]$ | $8.67 \times 10^{-18}$ |

Full CCL attains the highest terminal-window success rate under the frozen configurations. Its advantages over Static-All Restoration and DirectGate-Jacobian remain significant after Holm correction, as do those over the remaining handlers. These exploratory comparisons support higher completion under the evaluated configurations.

On the common-success subsets, the mean normalized-intervention differences (Full CCL minus comparator) are $-0.1586$ for DirectGate-Jacobian ($n = 41$, 95% CI $[-0.2114, -0.1054]$) and $-0.1010$ for Static-All Restoration ($n = 43$, 95% CI $[-0.1514, -0.0509]$). The TR-SQP common-success subset contains only two seeds ($-0.6108$, 95% CI $[-0.7247, -0.4969]$), and neither zero-completion handler has a common-success subset. These post-treatment comparisons are descriptive and do not estimate unconditional handler effects.

Static-All Restoration is the all-row restoration handler used in this benchmark, not the lifecycle-preserving row-exposure diagnostic in Section D.2. Reversible-Dwell-Hysteresis is the benchmark-specific reversible lifecycle baseline and is distinct from the supported-hit dwell extension evaluated in Section D.9.

### D.6 Controlled Late-Conflict Case

The controlled late-conflict case visualizes carry-forward protection in decoded video when an admitted commitment is challenged by a conflicting update. At $384 \times 576$ resolution, we evaluate the same prespecified candidate update in an unprotected branch, CCL–ExactSym, and CCL; Figure 3 shows the resulting outputs.

At the prespecified display frame 24, the Euclidean target-position errors in decoded pixels are $47.29$, $0.29$, and $0.28$ px under SAM2 (Ravi et al., 2025) for the unprotected, CCL–ExactSym, and CCL outputs, respectively. CoTracker (Karaev et al., 2024) gives $51.79$, $4.29$, and $4.29$ px in the same order. Both trackers cover all 49 frames, and measured background camera motion remains below $7 \times 10^{-4}$ in normalized coordinates.

This construction isolates protection under an imposed conflict. It does not estimate how often comparable conflicts occur in ordinary rollouts or support a perceptual ordering of CCL and CCL–ExactSym.

### D.7 Supplementary Matched-Satisfaction Results

Seed-wise ordering shows how consistently CCL reduces collateral deviation across seeds. CCL, Scalar, and CCL–ExactSym each satisfy the common final proxy criterion on all 100 paired seeds. Table 10 reports the seed-wise ordering together with the supplementary relative-effect intervals and sign-flip tests defined in Section C.

Table 10: **Supplementary matched-satisfaction results on 100 paired seeds.** Wins are reported as CCL/comparator. Intervals are nominal 95% paired-bootstrap CIs for the median per-seed relative reduction. The $p$-values are from two-sided paired sign-flip tests of the mean difference and are Holm-adjusted within each four-metric comparator family.

| Comparator | Metric | Wins (CCL/comparator) | Median reduction [95% CI] | Holm-adjusted $p$ |
|---|---|---|---|---|
| Scalar | Pixel | 100/0 | $82.21\%, [81.18\%, 83.27\%]$ | $4.00 \times 10^{-6}$ |
| | Latent | 100/0 | $86.83\%, [86.46\%, 87.19\%]$ | $4.00 \times 10^{-6}$ |
| | LPIPS | 100/0 | $88.56\%, [87.20\%, 90.16\%]$ | $4.00 \times 10^{-6}$ |
| | DINO | 100/0 | $89.38\%, [87.62\%, 90.63\%]$ | $4.00 \times 10^{-6}$ |
| CCL–ExactSym | Pixel | 84/16 | $24.73\%, [18.46\%, 29.85\%]$ | $4.00 \times 10^{-6}$ |
| | Latent | 83/17 | $20.86\%, [14.31\%, 25.57\%]$ | $4.00 \times 10^{-6}$ |
| | LPIPS | 97/3 | $29.64\%, [25.00\%, 35.42\%]$ | $4.00 \times 10^{-6}$ |
| | DINO | 87/13 | $27.91\%, [23.60\%, 33.40\%]$ | $4.00 \times 10^{-6}$ |

Against Scalar, all four relative-effect intervals exclude zero and all four Holm-adjusted tests remain significant. Against CCL–ExactSym, median per-seed reductions range from 20.86% to 29.64%, and all four mean-difference tests remain significant after Holm correction. The intervals and tests target different estimands, as defined in Section C.

### D.8 Parabolic Target-Family Replication

We repeat the matched-satisfaction comparison under a different calibrated target family. All three controlled methods satisfy the final proxy criterion on all 60 seeds.

Table 11: **Parabolic target-family replication on 60 paired seeds.** Each arrow reports the comparator median $\rightarrow$ the CCL median. Median reductions are medians of the per-seed relative reductions defined in Eq. (18). Mean differences are comparator minus CCL, so positive values favor CCL; intervals are nominal 95% paired seed-level percentile-bootstrap confidence intervals.

| Comparator | Metric | Median deviation | Median reduction | Mean difference [95% CI] |
|---|---|---|---|---|
| Scalar | Pixel | $0.083625 \rightarrow 0.014988$ | 82.56% | $0.066732\,[0.063249, 0.069801]$ |
| | Latent | $97.510746 \rightarrow 12.539646$ | 86.96% | $81.951187\,[77.765102, 85.445376]$ |
| | LPIPS | $0.401439 \rightarrow 0.042604$ | 89.50% | $0.346168\,[0.329520, 0.359792]$ |
| | DINO | $0.289840 \rightarrow 0.029110$ | 89.98% | $0.250147\,[0.238040, 0.260806]$ |
| CCL–ExactSym | Pixel | $0.019769 \rightarrow 0.014988$ | 21.45% | $0.005370\,[0.004122, 0.006689]$ |
| | Latent | $18.389779 \rightarrow 12.539646$ | 22.16% | $4.683225\,[3.596217, 5.807191]$ |
| | LPIPS | $0.058258 \rightarrow 0.042604$ | 27.21% | $0.021284\,[0.015748, 0.027177]$ |
| | DINO | $0.039493 \rightarrow 0.029110$ | 22.00% | $0.011832\,[0.008318, 0.015715]$ |

Against Scalar, CCL has lower deviation on all 60 seeds under every metric. Against CCL–ExactSym, the CCL win/loss counts are 56/4, 54/6, 49/11, and 54/6 for pixel, latent, LPIPS, and DINO, respectively. All four paired mean-difference confidence intervals exclude zero in favor of CCL. The parabolic replication therefore supports lower average deviation relative to both comparators.

### D.9 Dwell-Confirmed Admission

Finally, the dwell-confirmation stress test evaluates whether requiring consecutive supported hits prevents a one-step pulse-induced crossing from becoming an irreversible commitment. The pulse induces such a crossing in 4 of the 64 paired tasks. Table 12 summarizes false admission under the pulse and acquisition in the corresponding clean branch.

Table 12: **Dwell-confirmed admission on 64 paired tasks.** False admissions are evaluated in the pulse branch; clean acquisition follows the criteria in Section C.5.

| Admission rule | Pulse false admissions | Clean acquisition |
|---|---|---|
| FirstHit ($K = 1$) | 4/64 | 64/64 |
| Dwell $K = 2$ | 0/64 | 64/64 |
| Dwell $K = 3$ | 0/64 | 64/64 |

For each comparison with FirstHit, four pairs exhibit a false admission under $K = 1$ only and none exhibits a false admission under the longer dwell only. Thus, $B_2 = B_3 = 0.0625$, with a one-sided $97.5\%$ Tango lower confidence bound of $0.00234$. Under the prespecified Tango procedure, both contrasts meet the Bonferroni-adjusted criterion for false-admission risk reduction.

Clean acquisition is 64/64 under all three rules. For both $K = 2$ and $K = 3$, the paired difference from $K = 1$ is zero, with a nominal $95\%$ interval of $[-0.05662, 0.05662]$. These clean-branch comparisons are descriptive because no noninferiority margin was prespecified. The prespecified smallest-passing rule selects $K^\star = 2$; the selection reflects parsimony rather than an observed advantage over $K = 3$, and it does not alter the FirstHit ($K = 1$) policy used in the earlier experiments.

This result is specific to the prescribed one-step pulse and does not establish robustness to arbitrary observer errors.

# E Mechanism Analyses

## E.1 Finite-Step Guard Activity

A separate 40-seed diagnostic of the primary CCL realization records one guard outcome on each of the 24 controlled transitions following the six-transition initial hold, yielding 960 seed–transition observations. The counts are descriptive; transitions from the same seed are repeated observations.

Table 13: **Finite-step guard activity.**

| Recorded quantity | Count |
|---|---|
| Full step, $\alpha = 1$ | 913 |
| Smaller positive step, $0 < \alpha < 1$ | 38 |
| Zero step, $\alpha = 0$ | 9 |
| Subtotal with $\alpha < 1$ | 47 |
| Post-guard admitted-gate violations | 0 |

The guard therefore selects $\alpha < 1$ on 47 of 960 transitions, including nine with $\alpha = 0$. These results show that the finite-step check is active in practice, but they do not identify why a full-step candidate fails or what endpoint would have resulted without the guard.

## E.2 Same-State Numerical Row-Expansion Audit

The held-out same-state analysis uses one prespecified state–proposal pair from each of eight held-out seeds (3040–3047), taken from the CCL rollout at $\tau = 6$, the first nontrivial protected transition. Within each pair, the state, raw sampler proposal, and lifecycle-induced commitment set are held fixed; only the row family and protection operator vary. The eight seed–state pairs are the resampling units. Intervals are nominal $95\%$ percentile-bootstrap confidence intervals for the mean contrast, based on 10,000 paired resamples.

Let P, PV, and PVA denote the position, position-plus-velocity, and position-plus-velocity-plus-acceleration row families, respectively. For each row family $R$, define

$$H_R = \frac{1}{2}\left(\|\Delta z - u_R^{\mathrm{ExactSym}}\|_2^2 - \|\Delta z - u_R^{\mathrm{CCL}}\|_2^2\right). \tag{19}$$

Positive $H_R$ indicates that CCL makes a smaller squared local intervention than CCL–ExactSym for row family $R$ at the same state and proposal. Under the fixed-$J$ comparison of Corollary 2, the corresponding headroom equals $\frac{1}{2}\|q_t\|_2^2$.

For comparisons across row families, let $d = \Delta z$ and define

$$\Delta_R^{\mathrm{eq}} = \|d - u_R^{\mathrm{ExactSym}}\|_2^2 - \|d - u_{\mathrm{P}}^{\mathrm{ExactSym}}\|_2^2,$$

and

$$\Delta_R^{\mathrm{one}} = \|d - u_R^{\mathrm{CCL}}\|_2^2 - \|d - u_{\mathrm{P}}^{\mathrm{CCL}}\|_2^2, \qquad R \in \{\mathrm{PV}, \mathrm{PVA}\}.$$

The cross-family headroom contrast then satisfies the accounting identity

$$H_R - H_{\mathrm{P}} = \frac{1}{2}\left(\Delta_R^{\mathrm{eq}} - \Delta_R^{\mathrm{one}}\right).$$

For ideal projections, row expansion only adds one-sided constraints, so $\Delta_R^{\mathrm{one}} \geq 0$. This identity separates the change in the implemented equality projection from the additional intervention made by one-sided protection. It is an accounting decomposition, rather than an additional geometric guarantee.

Table 14: **Same-state numerical row-expansion contrasts under the finite-precision direct-row realization ($N = 8$ seed–state pairs).** Rows report implementation-level headroom contrasts relative to the position-only family. A positive value means that the computed headroom increases under this numerical pipeline; it is not an exact-arithmetic advantage of adding temporal rows.

| Headroom contrast | Mean | 95% CI | Pairs with positive contrast |
|---|---|---|---|
| $H_{\mathrm{PV}} - H_{\mathrm{P}}$ | 0.004904 | [0.004110, 0.005779] | 8/8 |
| $H_{\mathrm{PVA}} - H_{\mathrm{P}}$ | 0.009360 | [0.008114, 0.010640] | 8/8 |

Every exact structured row is a linear combination of the exposed position-error rows. Consequently, for exact, untruncated Jacobians, the equality row space is identical for P, PV, and PVA, whereas the corresponding one-sided feasible cones are nested:

$$\mathcal{C}_{\mathrm{PVA}} \subseteq \mathcal{C}_{\mathrm{PV}} \subseteq \mathcal{C}_{\mathrm{P}}.$$

The exact equality projection is therefore unchanged, while the distance from the proposal to the one-sided feasible cone cannot decrease. Hence,

$$H_{\mathrm{PVA}} \leq H_{\mathrm{PV}} \leq H_{\mathrm{P}}$$

for the exact, untruncated operators.

This ordering does not automatically extend to an implementation that applies a family-specific relative numerical-rank cutoff. Adding algebraically dependent rows can alter the singular spectrum and thereby change which directions survive the cutoff. Likewise, a factorized finite-precision realization is not equivalent to exact arithmetic when the numerical-rank rule is reapplied independently to each row family.

The positive contrasts in Table 14 therefore establish a narrower implementation-level observation: under the finite-precision direct-row pipeline, the computed headroom increases on all eight held-out pairs. The current table reports only the net contrast and does not, by itself, establish that off-span pseudo-modes cause this sign. Establishing that mechanism requires separately reporting the equality and one-sided terms above, the retained ranks and cutoff-adjacent singular values, and a common-projector factorized control.

This analysis is local, same-state, and pre-guard. It neither shows that temporal rows add independent exact first-order information nor establishes an endpoint advantage between independently evolving rollouts. It does not alter the fixed-$J$ decomposition in Corollary 2.

### E.3 Endpoint Row-Family Analysis

We next examine whether the CCL–ExactSym endpoint contrast changes as the protected row family is expanded from position alone.

For metric $x$ and row family $R$, define

$$A_R^x(s) = D_x(\mathrm{ExactSym}_R, s) - D_x(\mathrm{CCL}_R, s),$$

where $\mathrm{ExactSym}$ denotes CCL–ExactSym, and define

$$I_{R-\mathrm{P}}^x(s) = A_R^x(s) - A_{\mathrm{P}}^x(s), \qquad R \in \{\mathrm{PV}, \mathrm{PVA}\}.$$

Positive values indicate that expanding the row family increases CCL's endpoint-deviation advantage over CCL–ExactSym. Both methods satisfy the final gate on all 32 seeds under every row family, so each interaction uses the complete paired cohort. Intervals are nominal $95\%$ percentile-bootstrap confidence intervals for the mean interaction, based on 4,000 paired seed-level resamples.

Table 15: **Endpoint row-family interactions on 32 paired seeds.** Entries are mean interactions; positive values indicate a larger CCL endpoint-deviation advantage under the expanded row family.

| Interaction | Mean | 95% CI |
|---|---|---|
| PVA − P, pixel | $-0.000040$ | $[-0.003878, 0.003541]$ |
| PVA − P, latent | $0.070574$ | $[-4.992070, 4.907770]$ |
| PV − P, pixel | $-0.000564$ | $[-0.004111, 0.003146]$ |
| PV − P, latent | $-0.578332$ | $[-5.324110, 4.095430]$ |

All four intervals include zero, so this analysis does not resolve whether the endpoint CCL–ExactSym contrast changes with the protected row family. This is compatible with the same-state result because the two analyses target different estimands: pre-guard operator contrasts at matched states and proposals versus endpoints of independently evolving rollouts.

### E.4 Aggregate-Construction Stress Test

The per-frame construction retains separate lifecycle cells for the framewise residuals. The aggregate construction instead uses a single lifecycle cell with residual

$$c_{\mathrm{agg}}(z) = \max_k c_k(z)$$

and its associated Jacobian row. These constructions change numerical rank, lifecycle granularity, and residual aggregation simultaneously; the comparison therefore does not isolate rank.

For metric $x$ and construction $m \in \{\mathrm{per\text{-}frame}, \mathrm{aggregate}\}$, let

$$A_m^x(s) = D_x(\mathrm{ExactSym}_m, s) - D_x(\mathrm{CCL}_m, s), \qquad I^x(s) = A_{\mathrm{per\text{-}frame}}^x(s) - A_{\mathrm{aggregate}}^x(s).$$

A positive interaction indicates a larger CCL endpoint-deviation advantage under the per-frame construction. CCL and CCL–ExactSym satisfy the final gate under both constructions on all 60 seeds, so the analysis uses 60 complete four-condition pairs. Intervals are nominal $95\%$ percentile-bootstrap confidence intervals for the mean interaction, based on 4,000 paired seed-level resamples.

Table 16: **Per-frame versus aggregate endpoint stress on 60 paired seeds.** Advantage columns report medians of $A_m^x$; the final column gives a confidence interval for the mean per-seed interaction $I^x$.

| Metric | Per-frame median advantage | Aggregate median advantage | Mean-interaction 95% CI |
|---|---|---|---|
| Pixel | $0.001848$ | $0$ | $[-0.001941, 0.003959]$ |
| Latent | $1.39040$ | $0$ | $[-4.16296, 3.50429]$ |

Both mean-interaction intervals include zero. Because the constructions differ in three respects, this experiment neither isolates nor rules out a contribution from numerical rank.

# F TRANSFER, COST, AND SCOPE ANALYSES

This appendix extends the main evaluation along three dimensions. We first test the matched-satisfaction comparison on a second video-diffusion backbone. We then report the computational cost of the primary implementation and examine lower-cost Jacobian realizations and a lighter control interface. Finally, we assess sensitivity to the observer and target family and examine the referential, task-level, decoded-output, and perceptual scope of the numerical proxy. Because these studies use distinct cohorts and estimands, we report them separately.

## F.1 LTX-VIDEO TRANSFER

We evaluate the matched-satisfaction comparison under a fixed LTX-Video (HaCohen et al., 2024) 2B configuration. The experiment uses $256 \times 384$ decoded video, 33 decoded frames, 30 denoising steps, and 32 held-out seeds (9200–9231). It retains the mean-channel spatial-softmax centroid and linear target family used in the primary experiment but uses a different calibration point. After six shared warm-start steps, each seed-specific target is calibrated using the observer at the terminal state of the shared free warm start, evaluated at $\sigma = 0$, and is then held fixed across arms for the remaining 24 controlled transitions.

The controlled interface comprises 33 position gates and 192 outward structured relations: 66 position, 64 velocity, and 62 acceleration rows. The protected arms use directly differentiated structured rows; CCL solves its one-sided projection matrix-free. All 32 seeds are fully paired, and every controlled arm satisfies the matched final gate.

For comparator $b$, seed $s$, and distance metric $x$, define

$$\Delta_x^b(s) = D_x(b, s) - D_x(\mathrm{CCL}, s), \qquad \rho_x^b(s) = 100 \frac{\Delta_x^b(s)}{D_x(b, s)},$$

where $D_x$ measures final-output deviation from the corresponding aligned-free output. Positive values indicate lower deviation under CCL. Relative reductions are evaluated only for matched seeds with $D_x(b, s) > 10^{-12}$; mean differences use all 32 matched seeds. Table 17 reports the median per-seed relative reduction $\rho_x^b$ and a paired-bootstrap interval for the mean raw-distance difference $\Delta_x^b$.

Table 17: **LTX-Video transfer on 32 paired seeds.** All controlled arms satisfy the matched final gate on 32/32 seeds. Reductions are medians of per-seed relative reductions. Intervals are nominal 95% percentile-bootstrap confidence intervals for the mean paired raw-distance difference, computed from 10,000 seed-level resamples. Positive differences favor CCL.

| | Pixel | | Latent | |
|---|---|---|---|---|
| Comparator | Median reduction | 95% CI for mean $\Delta$ | Median reduction | 95% CI for mean $\Delta$ |
| Scalar | 85.66% | $[0.23062, 0.26781]$ | 85.66% | $[151.162, 178.706]$ |
| CCL–ExactSym | 41.50% | $[0.01454, 0.06294]$ | 15.45% | $[-1.941, 30.073]$ |

Both intervals exclude zero in favor of CCL over Scalar. Relative to CCL–ExactSym, the pixel interval excludes zero, whereas the latent interval does not. The experiment therefore reproduces the lower final-output deviation relative to scalar rescaling under this fixed LTX-Video configuration, but it does not establish a backbone-independent ordering between one-sided and equality protection.

## F.2 RUNTIME AND MEMORY

Table 18 reports median runtime and median peak CUDA allocation for the CogVideoX-2b trajectory-proxy implementation across 16 matched seeds. Runtime ratios compare each arm's median runtime with that of the aligned-free rollout. Costs for the Tora object-trajectory study are reported separately in Appendix G.9.

The primary direct-row CCL implementation is an auditable reference realization rather than a throughput-oriented controller. Its median runtime is 1243.72 s, or 38.95 times that of the aligned-

Table 18: Runtime and memory on 16 matched seeds. Ratios are computed from median runtimes, using the aligned-free rollout as the reference. The final column reports the median peak CUDA allocation across seeds.

| Arm | Median runtime | Runtime ratio | Median peak allocation |
|---|---|---|---|
| Aligned-free rollout | 31.93 s | 1.00× | 5.93 GiB |
| Scalar | 167.20 s | 5.24× | 5.87 GiB |
| CCL–ExactSym | 1239.16 s | 38.81× | 27.33 GiB |
| CCL | 1243.72 s | 38.95× | 27.37 GiB |

free rollout. Its median peak CUDA allocation is 27.37 GiB, compared with 5.93 GiB for aligned-free generation. The similar costs of CCL and CCL–ExactSym are consistent with their shared decoded observer, structured Jacobian, nonlinear guard, and controlled rollout. We do not claim that the primary implementation is deployment-ready, and these aggregate measurements do not isolate component-level costs.

**Single-seed Jacobian-realization timing.** We compare three Jacobian realizations under otherwise fixed conditions. All three measurements use seed 0, CogVideoX-2b, and the same 30-transition schedule: a six-transition initial hold followed by 24 denoiser-proposal transitions. They also share the observer evaluated on VAE-decoded frames, the calibrated target, the 186 structured relations, the commitment lifecycle, and the nonlinear guard. All were obtained on the same RTX 6000 Pro. The timing scope is restricted to the controlled rollout under the shared seed setup; the B0 diagnostic, final shared decoding, and auxiliary outputs are excluded.

Table 19: **Single-seed Jacobian-realization timing.** All rows use the same task, hardware, and timing scope. Ratios are the direct-row controlled-arm time divided by the corresponding variant time. They are descriptive single-seed timings, rather than end-to-end throughput estimates or evidence of effectiveness-preserving acceleration. These timings do not replace the 16-seed measurements in Table 18.

| Realization | Arm time | Direct/variant arm-time ratio | Jacobian used by projection |
|---|---|---|---|
| Direct structured rows | 918.14 s | 1.00× | current-state $J_t$ |
| Exact factorized + prefix | 300.12 s | 3.06× | current-state $B_t J_{e,t}$ |
| Interval-3 + Broyden | 137.16 s | 6.69× | approximate $B_t \widetilde{J}_{e,t}$ |

These $N = 1$ measurements identify Jacobian realization as a candidate optimization target, but they do not establish an accelerated primary configuration. No matched-seed equivalence analysis is reported for final satisfaction, persistence, collateral deviation, guard violations, or resource use. Exact factorization retains a current-state factorized Jacobian and the same one-sided objective, although it is numerically distinct from direct-row materialization. Interval-3+Broyden additionally approximates the current-state Jacobian between hard refreshes. The nonlinear guard remains fail-closed on successfully completed transitions, but exact-current-$J$ geometric statements do not apply between refreshes.

**Alternative low-cost proxy instantiation.** We next instantiate the lifecycle-and-guard design through a lighter control interface. This controller retains persistent commitments, post-update admission, and the nonlinear fail-closed guard, but uses a latent-mean observer and 64 position relations. It retains 31 of 32 temporal DCT modes per coordinate, giving 62 differentiated base coordinates. It refreshes the Jacobian every 32 transitions and uses Broyden updates between refreshes. The raw proposal is always checked by the nonlinear guard; after the first rejection, subsequent controlled transitions use projected updates.

We evaluate seeds 0–3 in two repetitions. All eight runs satisfy the final gate, and no transition takes a zero step. Across the post-warm-up measurements on seeds 1–3, the mean controlled-arm runtimes are 0.774 s and 0.773 s for the two repetitions, respectively.

This study changes the observer, temporal basis, protection family, Jacobian realization, and projection policy. It is therefore neither an acceleration result nor a speed–quality comparison for the primary CCL configuration. It shows only that persistent commitments and the nonlinear guard can be instantiated under a different, lower-cost proxy contract.

### F.3 Observer and Target Sensitivity

We next vary the observer and target family while retaining the same structured-row control interface. Both analyses use directly differentiated current-state rows, with a dense one-sided solve for CCL and the exact symmetric projection for CCL–ExactSym. The observer study contains all 16 matched seeds for each observer, and the target study contains all 20 matched seeds for each target family.

For metric $x$, comparator $b$, and seed $s$, let

$$r_x(b;s) = 100\left(1 - \frac{D_x(\mathrm{CCL}, s)}{D_x(b, s)}\right), \qquad \Delta_x(b;s) = D_x(b, s) - D_x(\mathrm{CCL}, s),$$

where $D_x$ is the final-output distance to the aligned-free counterpart. Relative reductions are evaluated only on matched seeds with $D_x(b, s) > 10^{-12}$, whereas mean differences use the complete matched set. The tables report the median of $r_x$ on the former set. Bracketed intervals are nominal 95% confidence intervals for the mean paired raw-distance difference $\Delta_x$, obtained from 4,000 nonparametric seed-level bootstrap resamples using percentile endpoints. Positive values favor CCL.

Table 20: **Observer sensitivity.** Each observer is evaluated on 16 matched seeds. Entries against Scalar are median pixel/latent relative reductions. Entries against CCL–ExactSym give the median relative reduction followed by the nominal 95% confidence interval for the mean paired raw-distance difference.

| Observer | vs. Scalar pixel / latent | vs. CCL–ExactSym pixel [95% CI] | vs. CCL–ExactSym latent [95% CI] |
|---|---|---|---|
| Mean RGB | 82.66% / 87.04% | $3.24\%$ $[-0.00346, 0.00402]$ | $-10.19\%$ $[-6.532, 3.146]$ |
| Red saliency | 89.75% / 91.94% | $14.21\%$ $[-0.00019, 0.00344]$ | $17.54\%$ $[0.197, 4.780]$ |
| Red mass | 90.08% / 92.86% | $32.82\%$ $[0.00192, 0.00311]$ | $35.86\%$ $[2.367, 3.128]$ |

The median reductions relative to Scalar remain large across all three observers. The relationship with CCL–ExactSym is less uniform: both mean-RGB intervals include zero, the red-saliency interval excludes zero only in latent space, and both red-mass intervals exclude zero in favor of CCL. Thus, this analysis does not establish an observer-independent ordering between the two protected operators.

Table 21: **Target-family sensitivity.** Each target family is evaluated on 20 matched seeds. Entries follow the conventions of Table 20; bracketed intervals are nominal 95% percentile-bootstrap confidence intervals for the mean paired raw-distance difference.

| Target | vs. Scalar pixel / latent | vs. CCL–ExactSym pixel [95% CI] | vs. CCL–ExactSym latent [95% CI] |
|---|---|---|---|
| Hold | 82.68% / 86.99% | $-2.99\%$ $[-0.00467, 0.00350]$ | $-26.59\%$ $[-8.230, 2.444]$ |
| Linear | 82.27% / 87.11% | $3.74\%$ $[-0.00082, 0.00591]$ | $-10.78\%$ $[-3.545, 4.952]$ |
| Parabolic | 82.93% / 86.95% | $11.90\%$ $[-0.00443, 0.00374]$ | $8.23\%$ $[-8.104, 3.071]$ |
| Preset linear | 82.19% / 86.99% | $13.93\%$ $[0.00108, 0.00540]$ | $11.55\%$ $[-0.779, 5.316]$ |

The median reductions relative to Scalar are positive across all four target families. In contrast, only the preset-linear pixel interval excludes zero in the comparison with CCL–ExactSym. The 60-seed parabolic experiment in Table 11 remains the primary target-family replication.

### F.4 Referenced-Object Bridge

The mean-channel centroid observer does not by itself establish that measured image-plane motion remains attached to the intended object. We therefore use the frozen referenced-object protocol in Appendix C.6. The study contains eight held-out seed pairs for a single red-balloon task, with a

Table 22: Completion on two separate numerical-proxy task suites. Risk differences are Full CCL minus Static-All, so positive values favor CCL. Intervals are two-sided 95% Tango paired efficient-score confidence intervals. The two suites are evaluated separately and are not pooled.

| Study | $N$ | Full CCL | Static-All | Risk difference [95% CI] |
|---|---|---|---|---|
| Original frozen suite | 32 | 32/32 | 32/32 | $0.0000\ [-0.1072,\ 0.1072]$ |
| Exploratory discriminative follow-up | 32 | 30/32 | 21/32 | $0.28125\ [0.09366,\ 0.46481]$ |

pair-specific first-frame Cutie reference shared between Neutral and Control. The seed pair, rather than the 49 frames within a pair, is the statistical unit.

For every pair and in both arms, the frozen object reference is supported on all 49 frames. The pre-action-baseline-adjusted Control-minus-Neutral $y$-centroid displacement over the prespecified active window is positive in all eight pairs, with a mean of $7.7367$ pixels and a range of $7.7286$ to $7.7488$ pixels. A separate, non-gating, method-blind identity review judged the tracked instance to be the target balloon in all eight pairs.

Together, full track support, signed displacement, and the identity review are consistent with the prescribed image-plane motion remaining attached to the fixed balloon specified by the fixed first-frame reference. This referent-specific diagnostic covers one frozen task; it neither compares CCL with another constraint handler nor certifies open-world semantic identity or three-dimensional physical motion.

### F.5 Numerical-Proxy Multitask Evaluation

We further evaluate task-level completion beyond fixed-task seed replications. The original study contains a prespecified suite of 32 numerical-proxy tasks. A separately frozen discriminative follow-up contains another 32 tasks. Each task is defined before either arm is run and is evaluated under Full CCL and Static-All; completion follows the study's frozen all-gate criterion. The task is the statistical unit, and both suites contain complete paired outcomes for all 32 tasks.

Full CCL completes all 32 tasks in the original frozen suite. Because Static-All also completes all 32 tasks, this result provides no evidence of a comparative advantage. In the discriminative follow-up, Full CCL completes 30/32 tasks, compared with 21/32 for Static-All, yielding a paired risk difference of 0.28125 (95% CI $[0.09366,\ 0.46481]$). This exploratory follow-up supports a completion advantage under its frozen numerical-proxy protocol.

The two suites were constructed under separate frozen protocols and are therefore analyzed without pooling. Both use the original $K = 1$ (FirstHit) admission rule; the later dwell-confirmed extension in Appendix D.9 was not used in these rollouts. Accordingly, these results characterize completion on the two frozen numerical-proxy task suites rather than performance on semantic tasks or acquisition of arbitrary external targets.

### F.6 Certification-Timing Diagnostic

A prespecified reversible-pulse diagnostic compares pre-update and post-update certification under FirstHit ($K = 1$) on 25 eligible paired seeds. Post-update certification satisfies the terminal criterion in 24/25 cases, compared with 25/25 for pre-update certification. The post-minus-pre difference is $-4$ percentage points, with a two-sided 95% confidence interval of $[-12, 0]$ percentage points.

The interval uses 10,000 paired seed-level percentile-bootstrap resamples generated with `random.Random(20260711)`. Each resample averages 25 paired outcome differences drawn with replacement. Endpoints are the sorted bootstrap values at zero-based indices 249 and 9749, without interpolation.

The prespecified criterion requires both terminal-success rates to be at least 90% and the lower interval endpoint to exceed $-5$ percentage points strictly. The latter uses the lower tail of the two-sided 95% interval, corresponding to a nominal one-sided level of 0.025. Both variants meet the success-rate requirement, but the lower bound of $-12$ percentage points fails the noninferiority requirement. This diagnostic therefore establishes neither noninferiority nor superiority of post-update certification.

Table 23: Five-rater blinded human evaluation. Counts are individual ratings across five raters and 32 non-catch items per comparison. Percentages are descriptive.

| Comparison | Indistinguishable | CCL preferred | Comparator preferred |
|---|---|---|---|
| CCL vs. CCL–ExactSym | 43 (26.9%) | 92 (57.5%) | 25 (15.6%) |
| CCL vs. Scalar | 27 (16.9%) | 116 (72.5%) | 17 (10.6%) |

We retain post-update admission as part of the lifecycle definition: an irreversible commitment is written only when its gate is satisfied on the realized state $z_{t-1}$, which is the state carried into the next transition. This is a semantic requirement of the commitment lifecycle, rather than a choice justified by a terminal-completion advantage. Terminal completion alone does not measure whether pre-update certification writes a commitment whose gate fails on the subsequently realized state. No conclusion about such false admissions follows from this diagnostic, and the experiment does not evaluate dwell-confirmed admission.

### F.7 Decoded-Video Diagnostics

Decoded-video measurements serve as auxiliary output diagnostics; none enters admission, protection, or certification. A targeted VBench (Huang et al., 2024b) analysis covers 240 videos across aligned free, Scalar, CCL–ExactSym, and CCL, evaluated on six dimensions. Dynamic degree is zero for every arm and is therefore not discriminative in this setting; the remaining dimensions do not identify a uniformly best controlled method.

Accordingly, these auxiliary decoded-video measurements do not establish a perceptual ordering among the controlled methods and do not extend the numerical proxy certificate to semantic or physical correctness.

### F.8 Blinded Human Evaluation

We additionally conduct a blinded evaluation of scene preservation at 384-pixel resolution. Five raters evaluated 32 non-catch items for each comparison, yielding 160 individual ratings per comparison. Raters were blinded to method identity, and every rater passed both catch trials. We report the individual ratings descriptively; because each item receives five ratings, the 160 responses are repeated judgments on 32 items rather than 160 independent observations.

Raters preferred CCL over CCL–ExactSym in 57.5% of judgments and over Scalar in 72.5%. These ratings favor CCL for scene preservation in both comparisons.

## G External Object-Trajectory Evaluation

This study compares Tora with and without CCL on object trajectories specified before generation. Tora supplies the trajectory-conditioned sampling proposals in both arms. CCL protects admitted constraints using differentiable proposal corrections and nonlinear object-based checks. The comparison therefore evaluates the complete protection module, including its admission rule, correction procedure, and finite-step acceptance test. We also compare commitment-conditioned correction with All-frame constraint exposure and report computational costs.

### G.1 Tasks and External Targets

The evaluation contains 16 scenes, with one linear and one curved trajectory per scene and two seeds per trajectory, giving 64 paired tasks. Four separate scenes are used for development. Tasks specify a fixed camera and a single reference object whose prescribed path remains within the image.

For each scene, the reference image and its human-verified object mask are resized to the generation resolution, $H = 480$ and $W = 720$. Let $p_0$ be the resulting mask centroid and $L = \min(H, W) =$

Table 24: Sampling configuration for the external object-trajectory study. Correction budgets apply to Tora+CCL and All-frame after the shared prefix.

| Quantity | Setting |
|---|---|
| Backbone | Tora SAT I2V, CogVideoX-5B-based weights |
| Model configuration | `cogvideox_5b_tora_i2v.yaml` |
| Inference configuration | `inference_sparse.yaml` |
| Output | 49 frames at $480 \times 720$, exported at 8 FPS |
| Latent temporal positions | 13 |
| Denoising schedule | 50 transitions: 38 shared, 12 controlled |
| Sampler | `VPSDEDPMPP2MSampler` |
| DynamicCFG | `scale=6`, `exp=5` |
| Network precision | BF16 |
| VAE scale factor | 0.7 |
| Fixed generated frames | `fixed_frames=0` |
| Admission | $K = 2$, with $E_f \equiv 1$ |
| Acquisition deadline | Controlled state $r = 6$ |
| Correction budget | At most three rounds per transition |

480. For video frame $f \in \{0, \dots, F-1\}$, with $F = 49$ and $u_f = f/(F-1)$, the target paths are

$$\begin{aligned} \tau_f^{\text{line}} &= p_0 + a u_f v, \\ \tau_f^{\text{arc}} &= p_0 + a u_f v + b \sin(\pi u_f) v_\perp, \end{aligned} \tag{20}$$

where $v$ and $v_\perp$ are orthogonal unit vectors. The displacement parameters are $a = 0.16L$ and $b = \pm 0.06L$. Both online and independent position checks use $\epsilon = 0.02L = 9.6$ pixels.

Path directions are selected using source-image geometry, requiring the translated object bounding box to remain inside the image. The reference image, object mask, prompt, target coordinates, and sampling seeds are fixed before generation. Generated rollouts, first-hit times, and subsequent regression opportunities are not used to construct or select the targets.

Tora receives 49 ordered trajectory points on its integer $256 \times 256$ input canvas. For $\tau_f = (\tau_{f,x}, \tau_{f,y})$, the input points and their induced pixel coordinates are

$$\begin{aligned} c_f &= \left( \left\lfloor \frac{256 \tau_{f,x}}{W} + \frac{1}{2} \right\rfloor, \left\lfloor \frac{256 \tau_{f,y}}{H} + \frac{1}{2} \right\rfloor \right), \\ \rho_f &= \left( \left\lfloor \frac{W c_{f,x}}{256} \right\rfloor, \left\lfloor \frac{H c_{f,y}}{256} \right\rfloor \right). \end{aligned} \tag{21}$$

The quantization error is recorded for each target. Online checks and independent evaluation use the original $\tau_f$ throughout. The reference image supplies the conditioning input and defines the object to be tracked; the generated frame $f = 0$ is evaluated against $\tau_0 = p_0$.

### G.2 Paired Sampling

Both arms use the official SAT image-to-video implementation of Tora (Zhang et al., 2025b), with the CogVideoX-5B-based I2V weights. The model loads `tora/i2v/mp_rank_00_model_states.pt`, with the accompanying `vae/3d-vae.pt` decoder and `t5-v1_1-xxl` text encoder. The backbone, text encoder, and decoder remain frozen. The reference image, prompt, trajectory, initial noise, and first 38 denoising transitions are shared within each pair. The complete sampler state is then copied to the two arms for the remaining 12 transitions. Trajectory conditioning remains active throughout generation.

We index the common entry by $r = 0$ and subsequent controlled states by $r = 1, \dots, 12$. These states continue the original 50-step schedule; the sampler history and guidance clock are not restarted. Video-frame index $f$ and denoising-state index $r$ refer to distinct time axes.

The Tora arm executes its native proposals and records shadow commitments using the same admission rule as Tora+CCL. Diagnostic readouts do not change its sampling updates or random-number stream.

Tora+CCL first checks each native proposal against its existing commitments and invokes a correction only when that proposal fails the check.

Each scheduled transition contains one native Tora proposal. When a corrected endpoint is accepted, the sampler retains the denoised prediction computed at the actual starting state and the random-number state returned by that native proposal. The corrected latent becomes the next sampling state. Trial corrections neither advance the sampler clock nor replace the stored denoised prediction with a projected latent.

### G.3 Object Readouts and Carry-Forward Protection

**Online object gates.** Cutie (Cheng et al., 2024) supplies the online object track, using `cutie-base-mega.pth` with `model=base`. Every readout starts from a fresh inference state initialized with the reference image and its object mask, followed by all 49 generated frames. Tracking state is not shared across candidates or arms, and the reference mask is not inserted into generated frame zero.

Let $q_f(z)$ be the tracked mask centroid in frame $f$ and $\chi_f(z)$ indicate that the same reference track has a nonempty mask. The online gate is

$$b_f^{\mathrm{on}}(z) = \mathbf{1}[\chi_f(z) = 1 \ \wedge \ \|q_f(z) - \tau_f\|_2 \le \epsilon] \,. \tag{22}$$

An empty mask fails the gate. Track support is an online observation; object identity and integrity are evaluated separately below.

The committed set and dwell counters start at $A_0 = \varnothing$ and $\ell_{f,0} = 0$. After each executed transition, the counters and committed set are updated as

$$\begin{aligned} \ell_{f,r} &= b_f^{\mathrm{on}}(z_r)\min\{2, \ell_{f,r-1} + 1\}, \\ A_r &= A_{r-1} \cup \{f : \ell_{f,r} = 2\}. \end{aligned} \tag{23}$$

Each admitted frame retains its first admission time $a_f$. Rejected trial states do not contribute to dwell confirmation.

**Differentiable direction corrections.** A frozen DINOv2 ViT-S/14 encoder (Oquab et al., 2024), `facebook/dinov2-small`, provides a reference-based soft position readout for constructing corrections. Reference and candidate frames are resized to $448 \times 672$ before feature extraction. Let $\widehat{e}_p$ denote the normalized feature at patch $p$, and let $m_p$ be the reference mask weight on that patch. The reference prototype and soft centroid are

$$\begin{aligned} v_{\mathrm{ref}} &= \mathrm{normalize}\left(\sum_p m_p \widehat{e}_p(I_{\mathrm{ref}})\right), \\ w_{f,p}(y) &= \mathrm{softmax}_p\left(\beta \left\langle \widehat{e}_p(D(y)_f), v_{\mathrm{ref}} \right\rangle\right), \\ q_f^S(y) &= \sum_p w_{f,p}(y) x_p, \end{aligned} \tag{24}$$

where $x_p$ is the patch center expressed in output pixel coordinates and $\beta = 20$ for all tasks. The reference prototype remains fixed throughout sampling.

The soft readout is computed in FP32, with gradients propagated through feature extraction and decoding to the candidate latent. Its decoder uses the same VAE weights in a separate FP32 instance; the decoder used for Cutie checks retains its fixed evaluation precision. Model weights are not updated.

For the current state $z$, let $\bar{z}$ be the native Tora proposal and $d = \bar{z} - z$. If every committed online gate passes at $\bar{z}$, the native proposal is accepted directly. Otherwise, define

$$g_f^S(y) = \frac{\|q_f^S(y) - \tau_f\|_2^2}{2L^2}, \qquad J_f(y) = \nabla_y g_f^S(y)^\top. \tag{25}$$

The local constraints contain one soft squared-position-error row per admitted frame. The structured position–velocity–acceleration family used in the primary proxy study is not used in this instantiation.

The correction keeps the native proposal $d$, starting state $z$, committed set $A$, and reference levels $\eta_f = g_f^S(z)$ fixed throughout the transition. Starting from $u^{(0)} = 0$, each round solves the local problem

$$\begin{aligned} u^{(k+1)} &= \arg\min_u \frac{1}{2}\|u - d\|_2^2, \\ \text{subject to} \quad & g_f^S(v_k) + J_f(v_k)\big(u - u^{(k)}\big) \le \eta_f, \qquad f \in A, \\ & v_k = z + u^{(k)}, \qquad k = 0, 1, 2. \end{aligned} \tag{26}$$

The first round reduces to the one-sided constraint $J_A(z)u \le 0$. Later rounds retain the affine offset associated with their trial expansion point. The quadratic programs are solved with OSQP (Stellato et al., 2020) in double precision (`float64`), using absolute and relative tolerances of $10^{-6}$ (`eps_abs=eps_rel=1e-6`), `max_iter=10000`, and `polishing=true`.

After each round, the candidate is converted to the submission precision and checked with Cutie. The first candidate satisfying all existing online commitments is executed. Subsequent linearization, when needed, uses the actual converted trial state. In subsequent rounds, $u^{(k)}$ denotes the displacement of that state from $z$. The soft readout determines correction directions; acceptance depends on the Cutie gates.

If the correction budget produces no accepted candidate, the run terminates and remains in the evaluation denominator. Nonfinite states and exact-zero submitted updates also terminate the run. Neither rejected candidates nor zero updates advance the dwell or preservation clock.

### G.4 Independent Object Evaluation

Independent evaluation covers the common entry and every executed denoising state, using all 49 frames of each decoded video. SAM 2-assisted annotations (Ravi et al., 2025) use `sam2_hiera_small.pt` with `configs/sam2/sam2_hiera_s.yaml`. Human annotators verify the annotations while blinded to method identity and online acceptance decisions. Disagreements are adjudicated. These annotations do not enter the controller, admission rule, or candidate selection.

For each video frame and denoising state, annotators assess the presence $P_{f,r}$, identity $I_{f,r}$, and integrity $Q_{f,r}$ of the specified reference object, together with its centroid $\widehat{q}_{f,r}$. Independent frame-level satisfaction is

$$b_{f,r}^{\mathrm{obj}} = P_{f,r} I_{f,r} Q_{f,r} \mathbf{1}[\|\widehat{q}_{f,r} - \tau_f\|_2 \le \epsilon]\,. \tag{27}$$

Disappearance, replacement, duplication, splitting, destructive deformation, and uncertain identity fail the corresponding object check. Ordinary changes in pose do not automatically count as integrity failures. Position is measured for the reference object; another instance is not substituted when correspondence is lost.

The fixed-camera requirement is evaluated separately on the final video. At least 20 static background points are selected outside the prescribed object sweep. Let $V_f$ contain the points that can be evaluated in frame $f$, and let $M_{\mathrm{bg}}$ be the number of selected points. Define

$$B_{\mathrm{final}} = \max_f \operatorname*{median}_{j \in V_f} \frac{\|p_{j,f} - p_{j,0}\|_2}{L}. \tag{28}$$

The background criterion requires $B_{\mathrm{final}} \le \delta_{\mathrm{bg}}$ and $|V_f|/M_{\mathrm{bg}} \ge \rho_{\mathrm{bg}}$ in every frame, with $\delta_{\mathrm{bg}} = 0.01$ and $\rho_{\mathrm{bg}} = 0.8$. Insufficient point coverage fails the background criterion. This check is a final task requirement and does not enter the online commitment gates.

### G.5 Acquisition and Persistence Endpoints

Let $U_s = 1$ when the common entry for task $s$ contains a valid reference object in every frame but does not yet satisfy the full target trajectory. The eligible cohort contains $N_U = \sum_s U_s = 55$ tasks. Let $C_s = 1$ indicate completion of all 12 controlled transitions, and let $B_s = 1$ indicate that the final background criterion passes. For incomplete runs, $Y_{\mathrm{obj}}$, $Y_{\mathrm{on}}$, $Y_{\mathrm{strict}}$, and $F_{\mathrm{final}}$ are set to zero without evaluating unexecuted states.

For $a \in \{\mathrm{on}, \mathrm{obj}\}$, define

$$G_r^a = \prod_{f=0}^{F-1} b_{f,r}^a, \qquad T_a = \min\{r \geq 2 : G_{r-1}^a = G_r^a = 1\}, \tag{29}$$

where $b_{f,r}^{\mathrm{on}} = b_f^{\mathrm{on}}(z_r)$. The acquisition time is infinite if no such executed states exist. It records the first qualifying acquisition and is not reset after a later regression.

Timely independent full acquisition is

$$A_{\mathrm{obj}} = U_s \mathbf{1}[T_{\mathrm{obj}} \leq 6]. \tag{30}$$

The primary endpoint additionally requires continued satisfaction, sampling completion, and final background validity:

$$Y_{\mathrm{obj}} = U_s C_s B_s \, \mathbf{1}[T_{\mathrm{obj}} \leq 6] \, \mathbf{1}\left[G_r^{\mathrm{obj}} = 1 \text{ for every } r = T_{\mathrm{obj}}, \ldots, 12\right]. \tag{31}$$

Every successful task therefore includes at least six executed denoising transitions after confirmed full acquisition. A regression after full acquisition fails this endpoint even if the target is recovered later. A local regression before full acquisition may recover without precluding primary success.

Independent final-video success is

$$F_{\mathrm{final}} = C_s B_s G_{12}^{\mathrm{obj}}. \tag{32}$$

This endpoint permits late acquisition or recovery after regression and does not require $U_s = 1$. Thus, $Y_{\mathrm{obj}} \leq A_{\mathrm{obj}}$ and $Y_{\mathrm{obj}} \leq F_{\mathrm{final}}$.

Online diagnostics use $A_{\mathrm{on}} = U_s \mathbf{1}[T_{\mathrm{on}} \leq 6]$ and

$$Y_{\mathrm{on}} = U_s C_s \mathbf{1}[T_{\mathrm{on}} \leq 6] \, \mathbf{1}\left[b_{f,r}^{\mathrm{on}} = 1 \text{ for every admitted } f \text{ and every } r \geq a_f\right]. \tag{33}$$

The independently verified strict commitment diagnostic is

$$Y_{\mathrm{strict}} = Y_{\mathrm{obj}} Y_{\mathrm{on}} \, \mathbf{1}\left[b_{f,r}^{\mathrm{obj}} = 1 \text{ for every admitted } f \text{ and every } r \geq a_f\right]. \tag{34}$$

It checks each frame from its own online admission time and is therefore stricter than the task-level primary endpoint.

All success rates use the full set of 64 tasks. Initially satisfied targets do not count as new acquisitions, and invalid entries, incomplete runs, and execution failures remain in the denominator.

### G.6 Statistical Analysis

The primary comparison concerns $Y_{\mathrm{obj}}$. Timely acquisition $A_{\mathrm{obj}}$ and final-video success $F_{\mathrm{final}}$ are the two key secondary endpoints. Online and strict commitment outcomes are diagnostic.

For any binary endpoint $Y$, the paired effect is

$$\Delta_Y = \frac{1}{16} \sum_{c=1}^{16} \frac{1}{4} \sum_{j=1}^{4} \left(Y_{c,j}^{\mathrm{Tora+CCL}} - Y_{c,j}^{\mathrm{Tora}}\right). \tag{35}$$

Each scene receives equal weight. Confidence intervals use 50,000 paired scene-bootstrap draws with random seed 20260908. Each draw resamples 16 complete scenes with replacement, retaining all four tasks and both methods within each selected scene. The reported 95% intervals are the 2.5th and 97.5th percentiles. They are individual intervals, rather than simultaneous intervals across endpoints.

Supplementary two-sided scene sign-flip tests use Holm correction over $Y_{\mathrm{obj}}$, $A_{\mathrm{obj}}$, and $F_{\mathrm{final}}$. The sign-flip test assumes exchangeability under within-scene method-label swaps. Trajectory-family and scene-level summaries describe coverage and heterogeneity.

### G.7 Results and Failure Analysis

Table 25 reports the independent task outcomes and online diagnostics. Tora+CCL improves timely acquisition, acquisition with sustained satisfaction, and final-video success. The paired primary

Table 25: External object-trajectory outcomes on 64 paired tasks. Differences are Tora+CCL minus Tora, in percentage points, with 95% paired scene-bootstrap confidence intervals. The lower block contains online and commitment diagnostics.

| Endpoint | Tora | Tora+CCL | Difference [95% CI] |
|---|---|---|---|
| Timely acquisition, $A_{\mathrm{obj}}$ | 38/64 | 49/64 | 17.2 [4.7, 29.7] |
| Primary success, $Y_{\mathrm{obj}}$ | 25/64 | 45/64 | 31.3 [18.8, 42.2] |
| Final success, $F_{\mathrm{final}}$ | 29/64 | 47/64 | 28.1 [14.1, 42.2] |
| Online acquisition, $A_{\mathrm{on}}$ | 39/64 | 50/64 | 17.2 [4.7, 29.7] |
| Online strict success, $Y_{\mathrm{on}}$ | 27/64 | 48/64 | 32.8 [18.8, 45.3] |
| Verified strict success, $Y_{\mathrm{strict}}$ | 23/64 | 45/64 | 34.4 [20.3, 46.9] |

Table 26: Mutually exclusive failure categories. Rows are assigned in the displayed order within each panel; each arm contains all 64 tasks.

| Outcome | Tora | Tora+CCL |
|---|---|---|
| *(a) Primary acquisition and persistence* | | |
| No timely independent full acquisition | 26 | 15 |
| Acquired, but sampling incomplete | 0 | 2 |
| Acquired and completed, but subsequent satisfaction fails | 12 | 1 |
| Other conditions pass, but final background fails | 1 | 1 |
| Primary success | 25 | 45 |
| *(b) Independent final-video evaluation* | | |
| Sampling incomplete | 0 | 2 |
| Final object invalid | 8 | 4 |
| Object valid, but full trajectory fails | 26 | 10 |
| Object and trajectory pass, but background fails | 1 | 1 |
| Final success | 29 | 47 |

difference is 31.3 percentage points, with a 95% confidence interval of 18.8–42.2. The Holm-adjusted sign-flip $p$-values are 0.00366 for the primary endpoint, 0.03125 for timely acquisition, and 0.00684 for final-video success.

For the primary endpoint, both arms succeed on 23 tasks, only Tora+CCL succeeds on 22, only Tora succeeds on two, and neither succeeds on 17. Among the 22 Tora+CCL-only successes, Tora misses the acquisition deadline on 13 tasks and regresses after full acquisition on nine. Tora+CCL misses the deadline on the two tasks where only Tora succeeds. These discordant pairs yield a net gain of 20 tasks, corresponding to 31.25 percentage points before rounding.

Table 26 separates acquisition, persistence, completion, and final-output failures. Tora completes all 64 runs, whereas Tora+CCL completes 62. The two incomplete Tora+CCL runs terminate after the correction budget yields no accepted candidate; both remain failures in the primary and final success rates. They are not classified as observed object losses.

Tora's 12 post-acquisition satisfaction failures in Table 26(a) comprise ten full-target regressions and two independent object-validity failures. The strict commitment diagnostic additionally rejects two Tora tasks with local regressions before full acquisition that later recover and satisfy the primary endpoint. This accounts for the difference between Tora's 25 primary successes and 23 strict successes. Tora+CCL has 45 successes under both criteria. The first local online commitment is reached on 60 tasks in each arm, consistent with their shared behavior before protection becomes active.

For linear trajectories, timely acquisition, primary success, and final success are respectively 16/32, 10/32, and 12/32 for Tora, compared with 22/32, 21/32, and 22/32 for Tora+CCL. For curved trajectories, the corresponding counts are 22/32, 15/32, and 17/32, compared with 27/32, 24/32, and 25/32. Primary success favors Tora+CCL in 12 scenes, ties in three, and favors Tora in one. Scene counts are provided in Table 27.

A valid reference object is present in the final output of 56/64 Tora tasks and 58/64 Tora+CCL tasks. The paired difference is 3.1 percentage points, with a 95% confidence interval of $[-9.4, 15.6]$. Thus,

Table 27: Scene-level counts for reproducing the paired analysis. Each entry is an ordered pair (Tora, Tora+CCL) among the four tasks in that scene. Valid denotes production of a final video with a valid reference object.

| Scene | $A_{\mathrm{on}}$ | $A_{\mathrm{obj}}$ | $Y_{\mathrm{on}}$ | $Y_{\mathrm{obj}}$ | $Y_{\mathrm{strict}}$ | $F_{\mathrm{final}}$ | Valid |
|---|---|---|---|---|---|---|---|
| S01 | (3,3) | (2,2) | (2,3) | (1,2) | (1,2) | (1,2) | (4,4) |
| S02 | (1,3) | (1,3) | (1,3) | (1,3) | (1,3) | (1,3) | (3,4) |
| S03 | (3,2) | (3,2) | (3,2) | (2,1) | (2,1) | (2,1) | (3,4) |
| S04 | (0,2) | (0,2) | (0,2) | (0,2) | (0,2) | (0,3) | (3,4) |
| S05 | (2,2) | (2,2) | (1,2) | (1,1) | (1,1) | (1,1) | (4,2) |
| S06 | (3,3) | (3,3) | (0,3) | (1,3) | (0,3) | (2,3) | (4,4) |
| S07 | (0,2) | (0,2) | (0,2) | (0,2) | (0,2) | (0,3) | (2,4) |
| S08 | (2,4) | (2,4) | (2,4) | (2,4) | (2,4) | (2,4) | (4,4) |
| S09 | (4,4) | (4,4) | (2,3) | (1,3) | (1,3) | (3,3) | (4,3) |
| S10 | (3,3) | (3,3) | (3,3) | (3,3) | (3,3) | (3,3) | (4,3) |
| S11 | (3,4) | (3,4) | (2,4) | (3,4) | (2,4) | (3,4) | (4,4) |
| S12 | (3,3) | (3,3) | (3,2) | (2,2) | (2,2) | (2,2) | (3,3) |
| S13 | (3,3) | (3,3) | (1,3) | (1,3) | (1,3) | (2,3) | (4,3) |
| S14 | (4,4) | (4,4) | (3,4) | (3,4) | (3,4) | (3,4) | (4,4) |
| S15 | (3,4) | (3,4) | (2,4) | (2,4) | (2,4) | (2,4) | (4,4) |
| S16 | (2,4) | (2,4) | (2,4) | (2,4) | (2,4) | (2,4) | (2,4) |

the improvement in joint trajectory success does not establish a separate improvement in object validity. The median maximum framewise trajectory error across all tasks is 3.25% of the image short side for Tora and 1.45% for Tora+CCL; invalid-object and incomplete runs are assigned infinite error for this ranking statistic.

The results support improved acquisition and sustained adherence on the specified external tasks. The Tora+CCL–Tora comparison concerns the complete CCL protection module and does not isolate the contribution of the surrogate geometry, finite corrections, or admission rule. The guard preserves admitted Cutie gates on successfully executed transitions; object-level performance is established by the independent annotations. The study does not establish arbitrary target reachability or compute-matched superiority.

### G.8 All-frame Constraint Exposure

The All-frame ablation uses the same 64 tasks and the sampling configuration in Table 24. It retains CCL's admission rule, nonlinear acceptance checks, soft readout, QP objective, and three-round correction limit. The change is the frame set used to construct the constraints in Eq. (26):

$$\mathcal{S} = \begin{cases} A, & \text{Tora+CCL,} \\ \mathcal{F} = \{0, \dots, 48\}, & \text{All-frame.} \end{cases}$$

Each row retains the reference level $\eta_f = g_f^S(z)$ defined at the actual starting state of the transition. These levels remain fixed across correction rounds. All other numerical settings follow Appendix G.3.

Both corrected arms check native proposals and corrected candidates against their own admitted Cutie gates. Correction is invoked when the native proposal fails this check. The two arms share the reference image, target, initial noise, and first 38 denoising transitions. The complete sampler state is copied at the common entry, after which each arm evolves its own sampling state, proposals, dwell counters, and commitment set over the remaining 12 transitions.

**Evaluation and statistical analysis.** Independent annotations and endpoint definitions follow Appendices G.4 and G.5. All success rates use the full denominator of 64 tasks, including the 55 tasks eligible for acquisition at the common entry. The paired differences are Tora+CCL minus All-frame, with equal weight assigned to each scene as in Eq. (35). Confidence intervals use 50,000 paired scene-bootstrap draws with seed 20260908, retaining the four tasks and both methods within each sampled scene. Two-sided sign-flip tests enumerate all $2^{16}$ scene-level sign assignments. Holm correction is applied across $Y_{\mathrm{obj}}$, $A_{\mathrm{obj}}$, and $F_{\mathrm{final}}$ within this comparison. The Tora+CCL–Tora comparison is reported separately in Appendix G.7.

Table 28: All-frame outcomes on 64 paired tasks. Differences are Tora+CCL minus All-frame in percentage points, with 95% paired scene-bootstrap confidence intervals. The endpoints are timely acquisition ($A_{\mathrm{obj}}$), primary acquisition and persistence ($Y_{\mathrm{obj}}$), and final-video success ($F_{\mathrm{final}}$). Holm correction covers these three endpoints.

| Endpoint | Tora+CCL | All-frame | Difference [95% CI] | $p$ | Holm $p$ |
|---|---|---|---|---|---|
| $A_{\mathrm{obj}}$ | 49/64 | 43/64 | $+9.4$ $[0.0, 18.8]$ | 0.15625 | 0.15625 |
| $Y_{\mathrm{obj}}$ | 45/64 | 32/64 | $+20.3$ $[9.4, 31.3]$ | 0.00781 | 0.02344 |
| $F_{\mathrm{final}}$ | 47/64 | 36/64 | $+17.2$ $[6.3, 28.1]$ | 0.02344 | 0.04688 |

**Results and failure analysis.** Tora+CCL achieves primary success on 45 tasks, compared with 32 for All-frame (Table 28). The paired difference is 20.3 percentage points, with a 95% confidence interval of 9.4–31.3. Primary and final-video success favor CCL after Holm correction at the 0.05 level; the timely-acquisition difference is not significant.

For the primary endpoint, both methods succeed on 30 tasks, only Tora+CCL succeeds on 15, only All-frame succeeds on two, and neither succeeds on 17. Among the 15 Tora+CCL-only successes, All-frame misses the acquisition deadline on six tasks, loses subsequent satisfaction on eight, and fails the final background criterion on one. On the two All-frame-only successes, Tora+CCL misses the deadline on one task and loses satisfaction after acquisition on the other. Table 26 reports the failure breakdown for Tora and Tora+CCL.

Both corrected arms complete 62 tasks. Their incomplete runs terminate after all three correction rounds fail the acceptance check. These runs remain in the primary and final-success denominators. Among completed All-frame outputs, five fail the object-validity check, 20 retain a valid object but fail the full trajectory, and one passes the object and trajectory checks but fails the background criterion.

Primary success favors Tora+CCL in ten scenes, favors All-frame in one, and ties in five (Table 29). For All-frame, timely acquisition, primary success, and final success are respectively 21/32, 14/32, and 16/32 on linear trajectories, and 22/32, 18/32, and 20/32 on curved trajectories. The corresponding Tora and Tora+CCL counts are reported in Appendix G.7.

Table 29: Scene-level All-frame outcomes. Each scene contains four tasks; $U$ counts entries eligible for acquisition. The middle three columns give All-frame endpoint counts. The last column counts tasks on which both All-frame and Tora+CCL meet the primary endpoint. Together with Table 27, these counts specify the scene-level paired comparisons.

| Scene | $U$ | $A_{\mathrm{obj}}$ | $Y_{\mathrm{obj}}$ | $F_{\mathrm{final}}$ | Both primary |
|---|---|---|---|---|---|
| S01 | 3 | 2 | 1 | 1 | 1 |
| S02 | 4 | 2 | 1 | 1 | 1 |
| S03 | 3 | 2 | 1 | 1 | 1 |
| S04 | 3 | 3 | 3 | 3 | 2 |
| S05 | 3 | 1 | 1 | 2 | 0 |
| S06 | 4 | 1 | 1 | 1 | 1 |
| S07 | 2 | 2 | 1 | 1 | 1 |
| S08 | 4 | 4 | 3 | 4 | 3 |
| S09 | 4 | 2 | 1 | 1 | 1 |
| S10 | 3 | 3 | 2 | 3 | 2 |
| S11 | 4 | 4 | 3 | 3 | 3 |
| S12 | 3 | 3 | 2 | 2 | 2 |
| S13 | 3 | 2 | 1 | 2 | 1 |
| S14 | 4 | 4 | 4 | 4 | 4 |
| S15 | 4 | 4 | 4 | 4 | 4 |
| S16 | 4 | 4 | 3 | 3 | 3 |
| Total | 55 | 43 | 32 | 36 | 30 |

**Constraint exposure.** Under CCL, 89 correction triggers occur while the admitted set is partial, $0 < |A| < 49$, spanning 44 tasks and 15 scenes. All 17 tasks with discordant primary outcomes lie

within this set. When $A = \mathcal{F}$, the two row-selection rules coincide. The results support using the commitment state to select correction constraints in this object-level implementation.

### G.9 Computational Cost

**Measurement.** All three methods use a single NVIDIA A100 80GB GPU with batch size one and the model, precision, and sampling settings in Table 24. Timing excludes model loading and warm-up. The common 38-transition prefix is shared across methods, and each method's suffix is accounted for separately. Full generation time for task $s$ and method $m$ is computed as

$$T^{\text{full}}_{s,m} = T^{\text{prefix}}_{s} + T^{\text{suffix}}_{s,m}.$$

Each method's full time includes the prefix cost once. The summaries include all 64 tasks, including runs that terminate before completing the suffix.

Generation time includes native sampling, control-related decoding and Cutie readouts, DINOv2 soft readouts and differentiation through the FP32 VAE, QP solving, admission and control processing, and output handling. Independent annotation and Tora's shadow-commitment diagnostics are accounted for separately. Component timings use non-overlapping scopes. Peak memory is the maximum GPU allocated memory for each task; both its mean and maximum across tasks are reported.

**Runtime and memory.** Tora+CCL takes 775.9 seconds per task on average, compared with 191.6 for Tora and 1014.3 for All-frame (Table 30). Its mean full generation time is 4.05 times that of Tora and 23.5% lower than All-frame. The corresponding suffix-time ratios relative to Tora are 13.01 for Tora+CCL and 17.91 for All-frame. The corrected arms have similar peak memory, averaging 43.8 and 43.9 GiB, compared with 35.5 GiB for Tora.

Table 30: Generation time and peak allocated memory on the 64 object tasks. Full times include the common prefix and each method's suffix. IQR gives the 25th–75th percentile interval, and P95 denotes the 95th percentile. Ratios in the lower panel compare mean full generation times; confidence intervals use paired scene bootstrap (Appendix G.6). All statistics are computed before rounding.

| Quantity | Tora | Tora+CCL | All-frame |
|---|---|---|---|
| Full time, mean (s) | 191.6 | 775.9 | 1014.3 |
| Full time, relative to Tora | 1.00× | 4.05× | 5.29× |
| Suffix time, mean (s) | 48.7 | 633.0 | 871.4 |
| Suffix time, relative to Tora | 1.00× | 13.01× | 17.91× |
| Full time, median (s) | 192.3 | 793.1 | 1010.6 |
| Full time, IQR (s) | [186.1, 197.7] | [462.2, 967.6] | [515.0, 1453.3] |
| Full time, P95 (s) | 201.3 | 1480.8 | 1897.2 |
| Peak memory, mean (GiB) | 35.5 | 43.8 | 43.9 |
| Peak memory, maximum (GiB) | 37.1 | 46.1 | 46.1 |
| Sampling completed | 64/64 | 62/64 | 62/64 |

| Mean-time ratio | Estimate | 95% CI |
|---|---|---|
| Tora+CCL / Tora | 4.050 | [3.669, 4.429] |
| All-frame / Tora | 5.294 | [4.687, 5.841] |
| Tora+CCL / All-frame | 0.765 | [0.712, 0.826] |

Time ratios are computed from the means across all 64 tasks before rounding. Their 95% confidence intervals use 50,000 paired scene-bootstrap draws following Appendix G.6, with the same scene draw shared across all three methods.

Table 31: Time components and correction workload. Component times are means in seconds per task and use non-overlapping timing scopes. Totals are computed before rounding. Workload counts are totals across all 64 tasks unless stated otherwise. Mean rows per QP round are weighted by the number of rounds. VJP counts refer to the row-wise differentiation used in this implementation.

| Component or workload | Tora | Tora+CCL | All-frame |
|---|---|---|---|
| *Time components: mean seconds per task* | | | |
| Shared 38-transition prefix | 142.9 | 142.9 | 142.9 |
| Native suffix sampling and output | 47.2 | 47.1 | 47.0 |
| Control decoding and Cutie readouts | 0.0 | 119.2 | 119.7 |
| Soft readouts and Jacobian construction | 0.0 | 460.2 | 697.3 |
| QP solving | 0.0 | 2.1 | 3.5 |
| Admission and other control processing | 0.0 | 2.9 | 2.9 |
| Scheduling and synchronization | 1.4 | 1.5 | 1.0 |
| Full generation time | 191.6 | 775.9 | 1014.3 |
| *Correction workload* | | | |
| Tasks invoking a QP | 0 | 50 | 50 |
| Correction triggers | 0 | 159 | 164 |
| QP rounds | 0 | 220 | 234 |
| Triggers using 1 / 2 / 3 rounds | — | 106 / 45 / 8 | 109 / 40 / 15 |
| Mean constraints per QP round | — | 34.1 | 49.0 |
| Row-wise VJPs | 0 | 7509 | 11466 |
| Online control Cutie readouts | 0 | 985 | 999 |
| Suffix transitions attempted | 768 | 765 | 765 |
| Suffix transitions executed | 768 | 763 | 763 |

**Sources of overhead.** Most of the time difference between the corrected arms occurs in soft readouts and Jacobian construction: 460.2 seconds per task for Tora+CCL versus 697.3 for All-frame (Table 31). QP solving accounts for only 2.1 and 3.5 seconds, respectively. The numbers of correction rounds are close, at 220 and 234, while the average numbers of constraints per round are 34.1 and 49.0. This corresponds to 7509 row-wise VJPs for CCL and 11466 for All-frame.

Accepted-candidate Cutie readouts are reused for admission. Consequently, the corrected arms use 985 and 999 online control readouts: one per native transition attempt plus one per correction round. The workload and timing results describe the object controller used in this evaluation.