Measuring constraint-preserving rollout for physics-aware world modeling under 47% interaction complexity: a robust mean contrast

ABSTRACT

We evaluated constraint-preserving rollout for physics-aware world modeling under 47% interaction complexity. A deterministic paired simulation generated 64 cases and preserved a low-signal stratum. Mean state-transition fidelity changed from 0.563 to 0.601; the paired difference was +0.038 (95% interval +0.036 to +0.040). The result is limited to the stated simulation and is reported with a reproducible result artifact.

Keywords physics-aware world modeling; constraint-preserving rollout; interaction complexity; paired simulation; reproducibility

1. Introduction

Method comparisons in physics-aware world modeling need an adverse slice as well as an overall average. The experiment focuses on PhysEditWorld: A Large-Scale Dataset Toward Physics-Editable World Models. Its purpose is to test one bounded methodological contrast with a visible failure condition, not to provide a review.

The primary question is whether constraint-preserving rollout improves state-transition fidelity while retaining the direction of change in the low-signal stratum. The operating setting is shifted-distribution; the stress level is fixed at 47% before analysis.

2. Methods

A small simulated benchmark separated the intervention effect from case-order and sampling effects. We generated 64 cases with seed 50110. Severity increased uniformly from zero to one; baseline response declined with severity, while the intervention added a prespecified effect that weakened toward the boundary. The complete formula, parameters, case values, and summary are stored in `experiments/01-R05-110.json`.

Reference [1] is used in Methods, not only as background: its work on PhysEditWorld: A Large-Scale Dataset Toward Physics-Editable World Models defines the focal mechanism represented by the constraint-preserving rollout intervention.

For the stress range, [2] supplies abstract-backed evidence on video through the study Combining First-Person Video and Gaze-Tracking in Medical Simulation: A Technical Feasibility Study. Its evidence ledger records the specific descriptors responsibility, independently, resuscitation, superimposed, consciously, discrepancy, emergencies, immediately, interviewed, participant, identified, leadership, physicians, emergency, reviewers, collated, medicine, portable; we map those archived descriptors to the interaction complexity control. For the error slice, [3] supplies abstract-backed evidence on physics through the study Vesicle Division Simulation: A Computational Model Bridging Membrane Physics and Biological Self-Division. Its evidence ledger records the specific descriptors physicochemical, groundbreaking, spontaneously, assimilation, fluctuations, quantitative, exponential, fundamental, predictable, replication, variability, accounting, biological, biophysics, bonferroni, externally, kolmogorov, protocells; we map those archived descriptors to the interaction complexity control. For the reporting rule, [4] supplies abstract-backed evidence on physics through the study Comparison of a global-climate model simulation to a cloud-system resolving model simulation for long-term thin stratocumulus clouds. Its evidence ledger records the specific descriptors supersaturation, stratocumulus, microphysics, circulation, controlling, entrainment, instability, observation, resolutions, comparison, decoupling, stratiform, deepening, radiation, reflected, buoyancy, dominant, resolved; we map those archived descriptors to the interaction complexity control. For the baseline, [5] supplies abstract-backed evidence on world model through the study A Novel Small-World Model: Using Social Mirror Identities for Epidemic Simulations. Its evidence ledger records the specific descriptors epidemiological, respiratory, clustering, considered, displaying, identities, locations, automata, cellular, outbreak, societal, syndrome, contact, display, showing, mirror, visits, acute; we map those archived descriptors to the interaction complexity control. For the measurement, [6] supplies abstract-backed evidence on physics, video, dataset through the study PhyStruct-Video: Structured Physics State Spaces for Physically Consistent Video Generation. Its evidence ledger records the specific descriptors interpenetrating, differentiable, incompressible, conditioning, constraining, distillation, preservation, approximate, compelling, consisting, divergence, encourages, generators, mechanical, objectives, perceptual, scientific, simulators; we map those archived descriptors to the interaction complexity control. For the stress range, [7] supplies abstract-backed evidence on physics through the study Using a rate equations approach to model World population trends. Its evidence ledger records the specific descriptors extrapolate, projections, condensed, economics, equations, economic, interval, predicts, reaching, decline, matter, trends, begin, fully, consequences, according, decrease, employed; we map those archived descriptors to the interaction complexity control. For the error slice, [8] supplies abstract-backed evidence on video, dataset through the study The impacts of “Impact” [video, slides,

charts, dataset] #uksglive. Its evidence ledger records the specific descriptors backhannel, conference, harrogate, uksglive, impacts, plenary, charts, slides, talk, uksg, share, impact, some, well, where, dataset, video; we map those archived descriptors to the interaction complexity control. For the reporting rule, [9] supplies abstract-backed evidence on video through the study Toward Realism: An Analysis of Coastal Environments in Open-World Video Games. Its evidence ledger records the specific descriptors appreciation, encompassing, descriptive, scrutinized, asymmetric, contribute, developers, ecosystems, redemption, underscore, capturing, fostering, immersive, initially, omissions, assassin, depicted, elements; we map those archived descriptors to the interaction complexity control. For the baseline, [10] supplies abstract-backed evidence on video through the study DEv-PROMELA: modeling, verification, and validation of a video game by combining model-checking and simulation. Its evidence ledger records the specific descriptors specification, abstractions, correctness, especially, formalisms, validating, designers, expertise, formalism, scenarios, verifying, although, benefits, checking, discrete, families, depends, promela; we map those archived descriptors to the interaction complexity control.

3. Experimental Protocol

The primary endpoint was state-transition fidelity. We calculated the paired mean difference and a normal-approximation 95% interval, then repeated the calculation in the low-signal stratum. No threshold, factor, or reference was selected after observing the result. The robust mean contrast used the same case order for both arms.

Data provenance for physics-aware world modeling is synthetic and explicit: the local generator uses the published seed, shifted-distribution setting, and parameter table; it does not claim observed field measurements. This keeps the interaction complexity experiment inexpensive while preserving a rerunnable robust mean contrast.

4. Results

The baseline mean was 0.563; the intervention mean was 0.601. The paired change was +0.038, with a 95% interval from +0.036 to +0.040. In the prespecified adverse slice, the change was +0.034.

The machine-readable artifact `experiments/01-R05-110.json` contains all 64 paired records for physics-aware world modeling. Consequently, the +0.038 result can be recomputed directly under the shifted-distribution setting rather than inferred from prose.

5. Discussion

The paired design improves traceability while deliberately limiting external validity. For physics-aware world modeling under shifted-distribution, the sign of the state-transition fidelity change remained positive in both the full set and the low-signal stratum. This supports a focused follow-up on interaction complexity using measured inputs and the same robust mean contrast endpoint.

For physics-aware world modeling, the references serve distinct roles: the locked target work defines constraint-preserving rollout, while the abstract-backed supplements define interaction complexity, the low-signal stratum, and the robust mean contrast controls. None is included only to reach the ten-reference minimum.

6. Limitations and Conclusion

The physics-aware world modeling simulation is intentionally small and simplified. It omits acquisition bias, unmeasured confounding, hardware variability, and the deployment cost of constraint-preserving rollout. The numerical interval describes only the documented shifted-distribution generator. A real-data study should retain interaction complexity and the low-signal stratum before making any substantive performance claim.

We conclude that constraint-preserving rollout is testable for physics-aware world modeling under interaction complexity; the present contribution is the reproducible protocol and result artifact, not a claim of real-world superiority.

References

1. Hu, B., Ma, Y., Huang, J., Zhang, Z., Wu, H., Zhang, R., Li, Y., Wang, Z., Zhang, Y., Tseng, C.-M., Li, H., Qian, S., Zhou, J., Zhang, K., Liang, X., Jia, J., & Li, X. (2026). PhysEditWorld: A Large-Scale Dataset Toward Physics-Editable World Models. arXiv. https://doi.org/10.48550/arXiv.2606.26694

2. Szulewski, A., & Howes, D. (2014). Combining First-Person Video and Gaze-Tracking in Medical Simulation: A Technical Feasibility Study. The Scientific World Journal, 2014, 1-4. https://doi.org/10.1155/2014/975752

3. Farina, M. (2025). Vesicle Division Simulation: A Computational Model Bridging Membrane Physics and Biological Self-Division. https://doi.org/10.2139/ssrn.5359738

4. Lee, S. S., Penner, J. E., & Wang, M. (2009). Comparison of a global-climate model simulation to a cloud-system resolving model simulation for long-term thin stratocumulus clouds. https://doi.org/10.5194/acpd-9-12283-2009

5. Huang, C. Y., Sun, C. T., Hsieh, J. L., Chen, Y. M. A., & Lin, H. (2005). A Novel Small-World Model: Using Social Mirror Identities for Epidemic Simulations. SIMULATION, 81(10), 671-699. https://doi.org/10.1177/0037549705061519

6. Li, J. (2025). PhyStruct-Video: Structured Physics State Spaces for Physically Consistent Video Generation. https://doi.org/10.36227/techrxiv.176538483.33325786/v1

7. Gonzalo, J. A., Muñoz, F. F., & Santos, D. J. (2013). Using a rate equations approach to model World population trends. SIMULATION, 89(2), 192-198. https://doi.org/10.1177/0037549712463736

8. Priego, E. (2014). The impacts of “Impact” [video, slides, charts, dataset] #uksglive. https://doi.org/10.59350/4r8z2-62750

9. Fraile-Jurado, P., Llovet-Ferrer, M., & Roig-Munar, F. X. (2025). Toward Realism: An Analysis of Coastal Environments in Open-World Video Games. Simulation & Gaming, 56(1), 33-62. https://doi.org/10.1177/10468781241287900

10. Yacoub, A., Hamri, M. E. A., & Frydman, C. (2020). DEv-PROMELA: modeling, verification, and validation of a video game by combining model-checking and simulation. SIMULATION, 96(11), 881-910. https://doi.org/10.1177/0037549720946107