SHORTCUT WORLD MODELS: LEARNING TO LEAP, NOT STEP

Anonymous authors
Paper under double-blind review

ABSTRACT

Autoregressive world models chain single-step predictions, requiring N forward passes for N steps into the future. We introduce Shortcut World Models, trained to predict environment dynamics across multiple horizons, enabling direct leaping to any learned step-size in a single pass rather than iteratively stepping through intermediate states. Beyond speed, skipping intermediate predictions also improves accuracy: errors compound through state discontinuities in autoregressive rollout, but shortcuts sidestep this accumulation entirely. At inference, adaptive chaining decomposes arbitrary horizons into learned sub-steps, handling step-sizes beyond training while maximizing accuracy with minimal sacrifice in speed. On discontinuous particle dynamics, Shortcut World Models achieve 33–64× fewer forward passes with up to 50% lower error, demonstrating a path toward learned simulators and model-based planning that are both faster and more accurate.

1 INTRODUCTION

World models predict future states from current state and action, and are central to model-based reinforcement learning and learned simulation (Ha & Schmidhuber, 2018; Hafner et al., 2025). The dominant paradigm is autoregressive: a model learns single-step transitions and chains predictions at inference to simulate multiple steps into the future. This is fundamentally limited. Each prediction feeds into the next, so small errors compound through state discontinuities, causing trajectory drift that grows with horizon length.

Recent work on one-step diffusion conditions on step size for fast sampling (Frans et al., 2024), with similar ideas appearing in latent-space video prediction (Hafner et al., 2025). For state-space dynamics with physical discontinuities, we decouple position along the trajectory (t, current time) from prediction horizon (Δt, step size), proposing **Shortcut World Models**: a model that predicts dynamics across multiple horizons directly, leaping to any trained step size in one forward pass.

A challenge with multi-horizon prediction is distribution shift: during chained inference the model receives its own predictions as input, but during training it only sees ground truth states. Venkatraman et al. (2015) formalize this and propose training on the model's own induced distribution, building on DAgger (Ross et al., 2011). We adopt this approach, teaching the model to recover from its own errors. At inference, adaptive chaining decomposes arbitrary target horizons into learned sub-steps, handling step sizes beyond training while maintaining accuracy.

We contribute: (1) Shortcut World Models, a horizon-conditioned architecture that predicts directly to any trained step size in one pass; (2) adaptive chaining, which decomposes arbitrary horizons into learned sub-steps; (3) empirical results on discontinuous particle dynamics showing 33–64× fewer forward passes and 8–50% lower error than autoregressive rollout.

2 ENVIRONMENT AND DATASET

We use a 2D particle simulator where a point mass moves in a bounded box $[-2\text{m}, 2\text{m}]^2$ with elastic wall collisions. The state $s = (x, y, v_x, v_y)$ represents position (m) and velocity (m/s); actions $a = (f_x, f_y)$ apply forces at each timestep. Actions are essential: without them, dynamics reduce to ballistic motion with predictable collision times. Any agent in a world model paradigm applies actions, so action-conditioned prediction is the relevant setting. The task is next-state prediction:

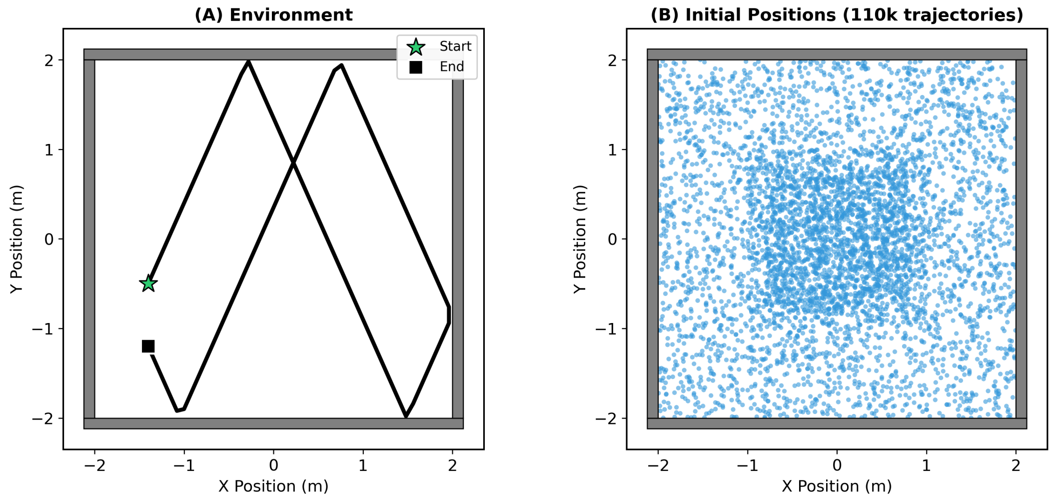

Figure 1: (A) Example trajectory with wall collisions. (B) Initial position distribution for 110k trajectories.

given current state (position, velocity) and action (force), predict the future state. This simulator serves as ground truth, using Euler integration between collisions and for training and evaluating both the sequential baseline and Shortcut World model.

Wall collisions create discontinuities: velocity reverses instantaneously without intermediate values. A small position error near a wall determines whether the model predicts a collision; missing one flips the velocity sign, causing predictions to diverge.We generate 100k training, 5k validation, and 5k test trajectories (85% have 4+ collisions). Each trajectory spans 1 second, simulated at step size $\Delta t = 0.01$ (101 timesteps).

3 SEQUENTIAL BASELINE

The sequential model predicts one step at a time. Given current state s_t and action a_t, the network outputs the predicted state change: $\hat{\Delta s} = f(s_t, a_t)$.

Architecture. An MLP with hidden layers [256, 256, 256, 128], LayerNorm, and GELU activations. Input dimension 6 (state + action), output dimension 4 (normalized deltas). Total: $\sim$105K parameters.

Training. MSE on normalized deltas: $\mathcal{L} = \|\hat{\Delta s} - \Delta s^{\text{norm}}\|^2$. Normalization balances position and velocity components which have different scales.

Inference. The next state is computed as $\hat{s}_{t+1} = s_t + \hat{\Delta s}$. To predict N steps, we chain N forward passes, feeding each predicted state as input:

$$\hat{s}_{t+k} = \hat{s}_{t+k-1} + f(\hat{s}_{t+k-1}, a_{t+k-1}), \quad k = 1, \ldots, N \tag{1}$$

Increasing capacity or using recurrent/attention architectures may help both approaches, but cannot eliminate the fundamental N-pass bottleneck of autoregressive rollout. We match architectures between Sequential and Shortcut to isolate the effect of horizon conditioning.

4 SHORTCUT WORLD MODELS

Unlike sequential models and neural ODEs (Chen et al., 2018), Shortcut World Models predict directly to any trained horizon in a single forward pass by conditioning on the step size. Given current state s_t, action sequence $\mathbf{a} = \{a_t, \ldots, a_{t+N-1}\}$, and step size Δt (where $\Delta t = N \times 0.01$), the model outputs:

$$\hat{s}_{t+N} = f(s_t, \mathbf{a}, \Delta t) \tag{2}$$

Architecture. Same MLP structure as Sequential (4 layers, [256, 256, 256, 128], LayerNorm, GELU). Input dimension is 205: state (4) + action buffer (100×2) + Δt (1). For horizon N, the first N action slots contain forces and the rest are zero-padded. Output dimension 4 (absolute state). Total: $\sim$411K parameters, larger due to action buffer, not added capacity.

Training. All step sizes $\Delta t \in \{0.01, 0.02, 0.04, 0.08, 0.16, 0.32, 0.64\}$ are trained simultaneously with ground truth supervision:

$$\mathcal{L}_{\text{sup}} = \mathbb{E}_{\Delta t}\left[\|f(s_t, \mathbf{a}, \Delta t) - s^{\text{GT}}_{t+N}\|^2\right] \tag{3}$$

(GT = ground truth). This differs from consistency models which use self-consistency losses; we supervise all horizons directly.

DAgger. Chained predictions at inference create distribution shift. We address this by training 10% of samples with DAgger: chain two predictions at half the step size and supervise against ground truth:

$$\hat{s}_{\text{mid}} = f(s_t, \mathbf{a}_{1:N/2}, \Delta t/2), \quad \hat{s}_{\text{final}} = f(\hat{s}_{\text{mid}}, \mathbf{a}_{N/2:N}, \Delta t/2) \tag{4}$$

$$\mathcal{L}_{\text{DAgger}} = \|\hat{s}_{\text{final}} - s^{\text{GT}}_{t+N}\|^2 \tag{5}$$

Gradients flow through both passes, teaching recovery from prediction errors. Total loss: $\mathcal{L} = 0.9 \cdot \mathcal{L}_{\text{sup}} + 0.1 \cdot \mathcal{L}_{\text{DAgger}}$.

Inference. For trained step sizes, a single forward pass suffices. For arbitrary horizons T, we decompose into trained steps via adaptive chaining. After training, we compute validation error for each Δt and define reliability $= 1/(\text{err} + \epsilon)$. Each step is scored by:

$$\text{score}(\Delta t) = \text{reliability}(\Delta t) + \frac{N}{T} \cdot \bar{r} \tag{6}$$

where $\bar{r}$ is mean reliability across all step sizes ($\epsilon = 10^{-6}$, computed once on validation set). We greedily select the highest-scoring Δt that fits the remaining horizon. For $T = 1.0$s, this yields $\{0.64, 0.32, 0.04\}$: 3 passes versus 100 for Sequential.

5 Results

We evaluate using position error (Euclidean distance to ground truth, averaged across 5000 test trajectories which are different from the validation or training dataset) and forward pass count.

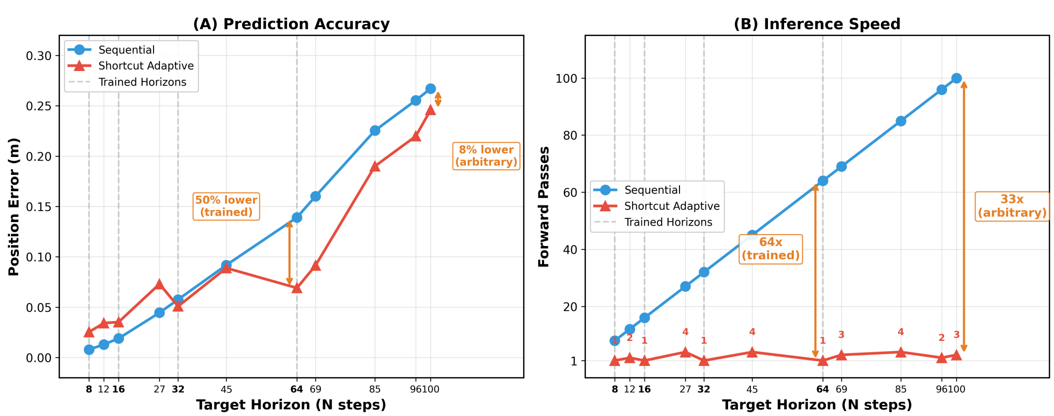

Figure 2: Position error and forward passes. Shortcut achieves 50% lower error at trained horizons (64x speedup) and 8% lower at arbitrary horizons (33x speedup).

Figure 2 compares Sequential and Shortcut Adaptive. At trained step sizes (N=64), Shortcut achieves 50% lower error with 64x fewer passes. At arbitrary horizons (N=100), adaptive chaining achieves 8% lower error with 33x speedup.

Why do certain step sizes have lower error? Particles collide every 0.3–0.7s. Step sizes like Δt=0.64 span one collision cleanly; smaller steps may land mid-collision where dynamics are most nonlinear. Step sizes should align with characteristic discontinuity timescales for a specific environment.

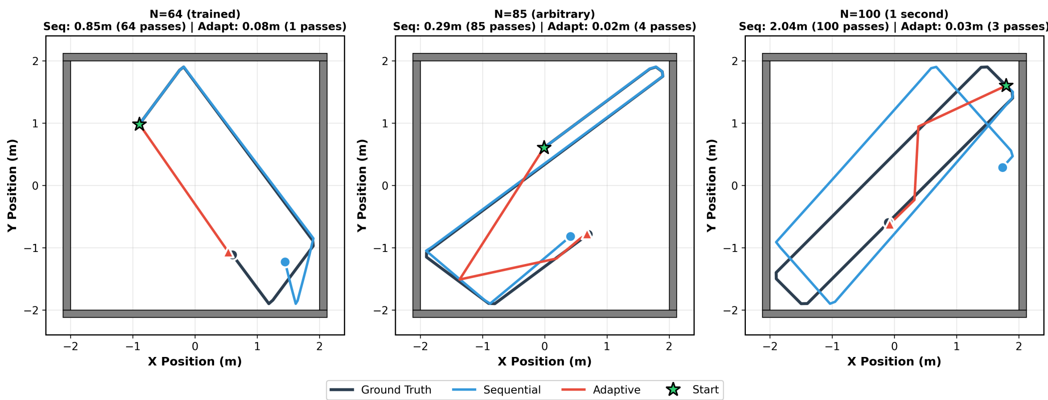

Figure 3: Trajectory comparison. Shortcut (red) matches ground truth (black); Sequential (blue) drifts.

Figure 3 shows trajectories with multiple collisions. At N=100, Sequential accumulates 2.04m error over 100 passes; Shortcut achieves 0.03m in 3 passes by jumping over discontinuities.

6 Conclusion

Shortcut World Models achieve 33 to 64x speedup with up to 50% lower error than autoregressive rollout. Since all step sizes are trained, dense intermediate states remain available when needed. This work is a proof of concept on a controlled domain; generalization across diverse environments with varying discontinuity patterns remains to be validated.

This opens exciting directions: (1) *surrogate simulators* for molecular dynamics and climate modeling where long-horizon prediction is bottlenecked by sequential rollout; (2) *adaptive MBRL* where fast multi-horizon rollouts let agents simulate thousands of action sequences and plan at variable temporal resolutions; (3) *scaling laws* for shortcut prediction, exploring whether larger models can learn longer horizons; (4) *architectural extensions* such as transformers or LSTMs with memory modules to capture longer temporal dependencies. We believe horizon-conditioned world models are a promising primitive for fast, flexible simulation.

References

Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. *Advances in neural information processing systems*, 31, 2018.

Kevin Frans, Danijar Hafner, Sergey Levine, and Pieter Abbeel. One step diffusion via shortcut models. *arXiv preprint arXiv:2410.12557*, 2024.

David Ha and Jürgen Schmidhuber. World models. *arXiv preprint arXiv:1803.10122*, 2018.

Danijar Hafner, Wilson Yan, and Timothy Lillicrap. Training agents inside of scalable world models. *arXiv preprint arXiv:2509.24527*, 2025.

Stéphane Ross, Geoffrey Gordon, and Drew Bagnell. A reduction of imitation learning and structured prediction to no-regret online learning. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pp. 627–635. JMLR Workshop and Conference Proceedings, 2011.

Arun Venkatraman, Martial Hebert, and J Andrew Bagnell. Improving multi-step prediction of learned time series models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 29, 2015.

7 Reproducibility and Ethics

Code and trained models will be released upon publication. This work poses no direct ethical concerns; however, faster simulation could accelerate both beneficial and harmful applications of autonomous agents.