MEND: MEASURING AND CONSTRAINING LATENT DYNAMICS IN WORLD MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

World models used for planning and simulation must preserve dynamics under repeated prediction. In joint-embedding predictive architectures (JEPAs), one-step accuracy does not ensure this behavior. We formulate latent prediction as an approximate semi-conjugacy. In the linear-symplectic setting, this relates dimensional collapse to encoder rank and prediction drift to predictor spectrum. MEND measures frequency, instability, and contraction in latent trajectories and uses these measurements to constrain a new predictor while keeping the encoder frozen. An orbit-averaged log-determinant separates physical dissipation from least-squares attenuation. An autonomously iterated V-JEPA 2 predictor has at least 12 positive finite-time Lyapunov exponents and loses an oscillator signal within 15 to 21 steps, while the encoder retains it. On a real pendulum video, MEND constrains a conformal-symplectic predictor in a two-dimensional observable. It extends the open-loop horizon from 21 to 651 steps (0.6 to 18.4 periods), approximately 31 times, while following the observed decay. These results support video-based system identification and prediction of mechanical motion using frozen representations. Extending the constraints to action-conditioned world models could support longer-horizon planning and control.

1 INTRODUCTION

Latent-space prediction reduces the cost of world modeling by evolving compact representations. This approach supports model-based control in PlaNet, Dreamer, DreamerV3, and TD-MPC2 (Hafner et al., 2019; 2020; 2025; Hansen et al., 2024). MuZero learns a planning model without reconstructing observations (Schrittwieser et al., 2020), while autoregressive video models predict visual tokens or learned changes in observation space (Wu et al., 2024; Ren et al., 2025). Feature-space world models advance the representations of a pretrained visual encoder (Zhou et al., 2025). JEPAs learn both the encoder and the predictor under joint-embedding prediction (LeCun et al., 2022; Assran et al., 2023; Bardes et al., 2024). Planning and simulation iterate the learned one-step map, so the predicted dynamics must remain accurate over the task horizon.

Given an observation x_t, an encoder E produces its latent representation, and a predictor P advances that representation by one step. Training makes the prediction match the target encoding of the next observation

$$P\big(E(x_t)\big) \;\approx\; \bar{E}(x_{t+1}), \tag{1}$$

where $\bar{E}$ is a stop-gradient or moving-average target encoder. This construction has two degenerate modes. In representation collapse, the objective (1) is minimized by a constant encoder, so the representation loses dimensions. In prediction drift, iterating P damps, displaces, or destabilizes the predicted trajectory even when the one-step error is small. Representation collapse removes state information, whereas prediction drift changes the evolution of the retained information.

Existing methods preserve latent variation through moving-average targets, stop-gradient, redundancy reduction, or moment matching (Grill et al., 2020; Chen & He, 2021; Caron et al., 2021; Zbontar et al., 2021; Bardes et al., 2022). LeJEPA's SIGReg shapes the latent distribution toward an isotropic Gaussian (Balestriero & LeCun, 2025). Spectral penalties pull the predictor inside the unit

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.

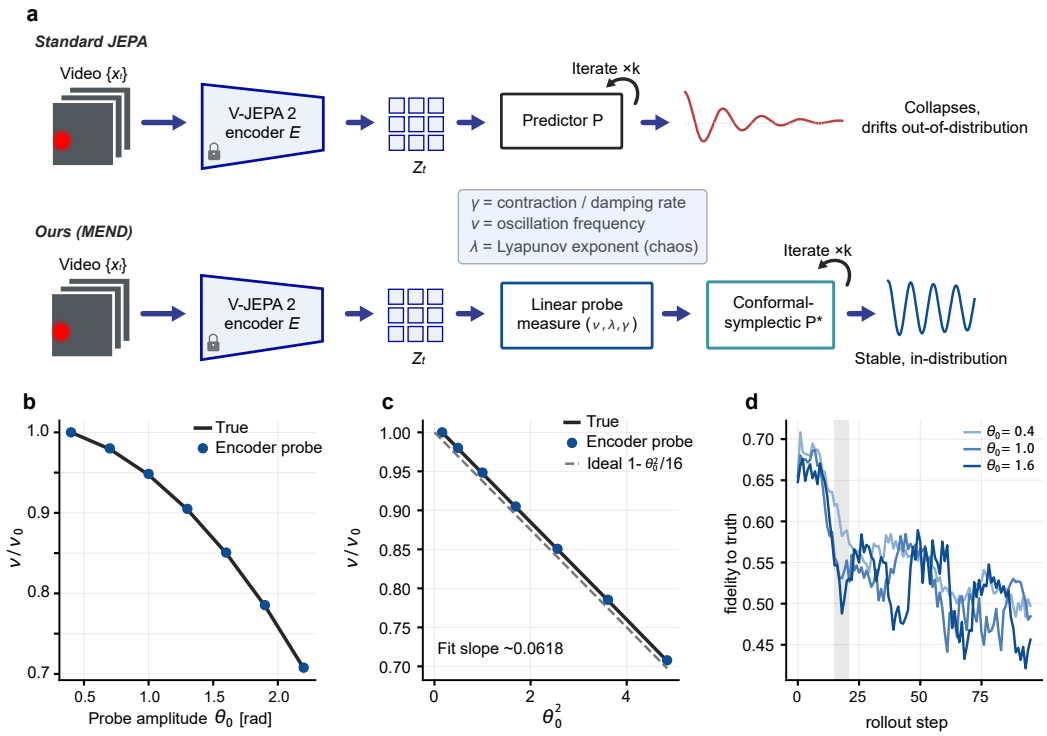


Figure 1: **(a)** Two prediction procedures sharing the same frozen V-JEPA 2 encoder E. **Standard JEPA (top):** autonomous iteration of P produces trajectories that drift out-of-distribution and collapse. The semi-conjugacy relation is $P \circ E \approx \bar{E} \circ \Phi$, with defect δ (see predictor attenuation and encoder-side nondegeneracy). **MEND (bottom):** the invariants (ν, λ, γ) are measured in the same representations and enforced in a conformal-symplectic predictor P^* that remains stable and in-distribution. **(b)** A linear probe of the frozen encoder recovers the pendulum's nonlinear dispersion to 0.01%. **(c)** The fitted tune-shift coefficient is -0.0618, compared with $-1/16$; $R^2 = 0.995$. **(d)** The iterated predictor loses the same signal within $\tau = 15$ to 21 steps at every amplitude.

circle (Lu et al., 2026), and receding-horizon planning limits exposure to open-loop drift (Hansen et al., 2024). Recent methods train direct long-horizon predictions or recovery from corrupted prediction histories (Li et al., 2026; Zhang et al., 2026). These methods keep the representation full-rank and the predicted trajectory bounded. A bounded trajectory can still damp conservative motion, so the spectrum alone does not separate fitted contraction from physical dissipation.

Distributional methods preserve latent variation without specifying the iterated dynamics (Balestriero & LeCun, 2025; Bardes et al., 2022; Maes et al., 2026; Zhao et al., 2026). Many structure-preserving predictors train representations jointly with decoders (Greydanus et al., 2019; Jin et al., 2020; Toth et al., 2019; Zhong et al., 2020; Chen et al., 2025). Operator-theoretic methods correct errors-in-variables attenuation in noisy linear dynamics (Takeishi et al., 2017; Lusch et al., 2018; Otto & Rowley, 2019; Azencot et al., 2020; Dawson et al., 2016; Hemati et al., 2017). Building on these results, the present analysis assigns collapse and drift to the encoder and predictor and distinguishes physical dissipation from artificial damping. Appendix C gives the extended discussion.

Here the JEPA relation is treated as an approximate semi-conjugacy to the true dynamics Φ. It describes how the predictor advances latent states as Φ advances phase-space states, exactly in the linear regime and up to a measurable defect otherwise. Collapse corresponds to a rank degeneracy of E, and prediction drift depends on the spectrum of the iterated P. This separation leads to **MEND**, the measure-and-enforce principle in Figure 1. Frequency, Lyapunov exponents, and contraction are measured in the representations of a frozen encoder and used to constrain the spectral class of a new predictor.

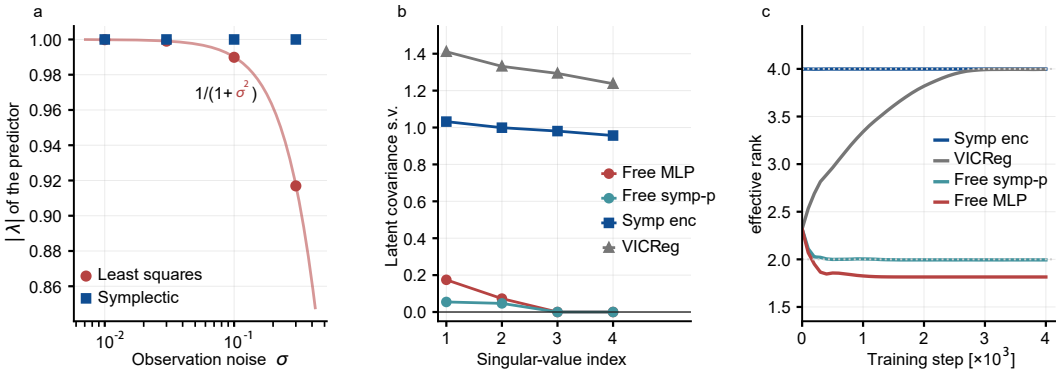


Figure 2: Tests of the structural predictions on the solvable system (Table A3). **(a)** The least-squares eigenvalue follows $1/(1+\sigma^2)$, with a maximum relative error of 5.5×10^{-4}; the symplectic fit stays on the unit circle. **(b)** Under the unregularized objective, the free encoder collapses to rank 2 with either predictor; a symplectic encoder or VICReg keeps all four dimensions. **(c)** During training, the free encoder's effective rank falls within a few thousand steps while the symplectic encoder holds rank 4.

The analysis yields three findings. First, a symplectic encoder prevents dimensional collapse, whereas a symplectic predictor does not constrain encoder rank. Second, observation noise attenuates a least-squares predictor by $1/(1+\sigma^2)$ (Dawson et al., 2016), so the one-step bias compounds geometrically under iteration. Third, the orbit-averaged log-determinant γ measures physical dissipation and is invariant under bounded-distortion changes of latent coordinates.

These measurements diagnose a frozen V-JEPA 2 model and specify its replacement predictor. The autonomous V-JEPA 2 predictor has at least 12 positive finite-time Lyapunov exponents and loses an injected oscillator after 15 to 21 steps. Both the pretrained and random-weight encoders retain the oscillator signal under the same probe, linking this decodability to visual observability. On one real pendulum video, a conformal-symplectic predictor supplied with the measured γ extends the open-loop horizon from 21 to 651 steps (0.6 to 18.4 periods). This improvement of approximately 31 times is achieved while following the observed decay.

2 Methods

Setup and training objective. Let $\Phi : \mathcal{X} \to \mathcal{X}$ denote the unknown dynamics on the observation space $\mathcal{X}$. A JEPA consists of an encoder $E : \mathcal{X} \to \mathcal{Z}$ into a latent space $\mathcal{Z} \subseteq \mathbb{R}^m$, a predictor $P : \mathcal{Z} \to \mathcal{Z}$, and a target encoder $\bar{E}$. Training minimizes the one-step prediction risk

$$\mathcal{L}(E, P) \;=\; \mathbb{E}_{x\sim\mathcal{D}} \left\| P(E(x)) - \bar{E}(\Phi(x)) \right\|^2 \quad , \tag{2}$$

where $\mathcal{D}$ is the data distribution over states. Starting from $z_0 = E(x_0)$, the n-step prediction is $P^{\circ n}(z_0)$.

Semi-conjugacy defect. The pair (E, P) is an exact semi-conjugacy for Φ if $E \circ \Phi = P \circ E$ on the support of $\mathcal{D}$. In general, $\delta(x) := \|P(E(x)) - E(\Phi(x))\|$ is the pointwise defect, and $\varepsilon^2 := \mathbb{E}_{\mathcal{D}}\delta(x)^2$ is the mean-square defect. Up to the target-encoder asymmetry, the JEPA objective (2) equals ε^2. The relation $E \circ \Phi \approx P \circ E$ is therefore enforced in $L^2(\mathcal{D})$.

Under repeated prediction, the predictor's geometry controls how the one-step defect accumulates. If $\delta \le \delta_0$ along an orbit and P is L-Lipschitz on a region containing both the predicted and encoded true orbits, Proposition A.3 gives

$$\|P^{\circ n}(E(x)) - E(\Phi^{\circ n}(x))\| \;\le\; \delta_0 \sum_{j=0}^{n-1} L^j \;=\; \begin{cases} n\delta_0, & L = 1, \\ \dfrac{L^n - 1}{L - 1}\delta_0, & L \neq 1. \end{cases} \tag{3}$$

For linear predictors, L is the induced operator norm; its relation to the spectrum depends on the choice of norm. An elliptic predictor has semisimple eigenvalues on the unit circle and achieves

$L = 1$ in an adapted norm (Theorem A.22). The prediction error then grows at most linearly with the horizon. A spectral radius above one amplifies the same defect exponentially. A spectral radius below one adds a systematic contraction bias (artificial damping, Theorem A.15). Long-horizon prediction therefore depends on both the spectrum of P and its one-step accuracy.

The theorems below assume linear symplectic dynamics and specify constraints on the encoder and predictor. The measurements ν, λ, and γ in §2 apply to differentiable latent predictors beyond this regime. They can be computed without ground-truth trajectories and used to diagnose a frozen model. Theorem A.24 and Lemma A.25 establish coordinate invariance for γ and the Lyapunov spectrum under their respective assumptions.

Dynamical readouts. Frequency-map analysis (Laskar, 1993) gives the dominant frequency ν and its drift $|\Delta\nu|$ between trajectory halves. The Benettin QR method gives the Lyapunov spectrum, with positive exponents identifying unstable latent directions. The singular values and log-determinant of $\mathrm{D}P$ measure local stretching and volume change. The orbit-averaged log-determinant γ is invariant under bounded-distortion reparametrization (Theorem A.24). The Lyapunov spectrum is invariant under bi-Lipschitz conjugacy (Lemma A.25), whereas ν and $|\Delta\nu|$ depend on the observable. Appendix E gives the operational definitions.

The measurements are first calibrated on systems with analytic answers. For the pendulum, calibration recovers the frequency to 7×10^{-11} and the $-1/16$ tune-shift coefficient. For the Chirikov standard map, it recovers the $K = 0$ rotation number to 3×10^{-9} and tracks the onset of chaos. The required one-step fidelity depends on the system, so each target model is calibrated separately.

The semi-conjugacy relation imposes separate conditions on the encoder and predictor. These conditions determine whether a latent propagator exists, whether the encoder can collapse, and how the predictor behaves under iteration. They also identify a contraction rate that is unchanged by admissible coordinate transformations. The linear statements use Assumption A.2; Theorem A.24 and Conjecture A.14 address the nonlinear setting. Appendix A gives the formal statements and proofs.

Lossless latent reduction. A latent propagator exists when $\ker E$ is invariant under the true dynamics. The reduction is lossless symplectic when $\ker E$ is a symplectic subspace. Canonical pairs (q_j, p_j) are therefore inseparable: whole (q, p) planes are kept or discarded. The minimal latent carrying k degrees of freedom has dimension exactly $2k$. An odd-dimensional latent space, or a representation containing a position without its conjugate momentum, cannot admit a lossless reduction.

Encoder-side nondegeneracy. The reduction must also remain nondegenerate. If the latent inherits a nondegenerate symplectic form from the encoder, then the encoder has full rank m. This rank condition excludes dimensional collapse without a stop-gradient, teacher-student asymmetry, or distributional regularizer. A symplectic predictor imposes no such constraint on the encoder and does not prevent collapse.

Predictor attenuation and stability. A nondegenerate encoder can still be paired with a predictor that distorts the dynamics. Consider $z_{t+1} = Rz_t$ with R orthogonal and symplectic, stationary state second moment $\mathbb{E}[z_t z_t^\top] = sI$, and observations $x_t = z_t + \varepsilon_t$ with independent, zero-mean isotropic noise of covariance $\sigma^2 I$. The population least-squares predictor is

$$\widehat{M}_\infty = \mathbb{E}[x_{t+1}x_t^\top]\,\mathbb{E}[x_t x_t^\top]^{-1} = \frac{R}{1+\sigma^2/s}. \tag{4}$$

The noise is independent of the states and across time, so the cross-moment remains sR while the input second moment increases to $(s+\sigma^2)I$. The fitted map therefore pulls every eigenvalue inward (Theorem A.15). After n steps, the predicted amplitude is reduced by $(1+\sigma^2/s)^{-n}$, although the true dynamics conserves amplitude. This is the classical errors-in-variables attenuation of noisy linear system identification (Dawson et al., 2016). A symplectic predictor excludes this uniform shrinkage because its spectrum is reciprocal. Volume preservation still permits hyperbolic growth; bounded prediction requires an elliptic spectrum with semisimple eigenvalues on the unit circle.

Coordinate-identifiable dissipation. Distinguishing this attenuation from physical dissipation requires a measure that is stable under changes of coordinates. For a C^1 diffeomorphism P and an

orbit $z_{t+1} = P(z_t)$, define

$$\widehat{\gamma}_n(z_0) = -\frac{1}{n\,m}\sum_{t=0}^{n-1}\log\bigl|\det \mathrm{D}P(z_t)\bigr|, \qquad \gamma = \lim_{n\to\infty}\widehat{\gamma}_n(z_0). \tag{5}$$

When the limit exists, it is invariant under every bounded-distortion change of latent coordinates. For a chart φ satisfying $\sup_z |\log|\det \mathrm{D}\varphi(z)|| \le K$, the finite-orbit readouts differ by at most $2K/(nm)$ (Theorem A.24). An unbounded-distortion chart can absorb dissipation entirely, so fixing the coordinates by whitening cannot measure it. The conformal-symplectic class admits dissipation at rate $c = 2\gamma$ while retaining the encoder-side nondegeneracy guarantee.

Under the assumptions of Lemma A.17, direct n-step regression gives the population contraction readout $\widehat{\gamma}_n = \gamma + \beta/(nm)$, where $\beta \ge 0$ is fixed across lags sharing the same starting-time window. The noise bias thus decays as $1/n$, whereas iterating the one-step fit retains the bias β/m at every horizon.

3 Results and Discussion

Figure 2 tests the structural implications on the solvable T_0 system, where encoder rank and predictor spectrum can be controlled independently. A symplectic predictor does not prevent collapse, whereas a symplectic encoder remains full rank without a distributional regularizer (Figure 2b). SIGReg and VICReg preserve rank but do not prevent the noisy least-squares spectrum from contracting to $1/(1+\sigma^2)$ (Figure 2a). The two constraint classes address different degenerate modes and can be combined. Appendix G reports the full comparison.

Figure 3 compares the information retained by the frozen V-JEPA 2 encoder with the dynamics generated by its predictor. The model contains a 1.0B-parameter ViT-g encoder and a 0.3B action-conditioned predictor (Assran et al., 2025). Fixed action and proprioceptive conditioning make the recurrent predictor an autonomous latent map. The encoder resolves an injected oscillator to approximately 1%, whereas the iterated predictor loses that signal within a few Lyapunov times.

All twelve estimated finite-time Lyapunov exponents are positive, ranging from $\lambda_1 \approx 0.15$ to 0.05 (Figure 3a). The estimates use Benettin QR over 130 iterations after a 20-step burn-in. Their signs and leading value remain stable across the tested finite-difference scales (Figure 3b). The exponents also remain positive across the tested initial conditions and context lengths. The leading exponent spans 0.04 to 0.16, corresponding to Lyapunov times of 6 to 23 steps. Appendix D gives the numerical settings and results for each context length.

Figure 1c shows the amplitude-dependent frequency recovered from self-rendered pendulum videos. A linear probe decodes the signed angle with $R^2 = 0.995$. Frequency-map analysis gives a leading coefficient of -0.0618, compared with the analytic $-1/16 = -0.0625$. Direct pixel tracking gives -0.059 ± 0.005 on real footage, compared with -0.060 ± 0.020 from the latent probe. The same pipeline applied to a randomly initialized ViT-g gives -0.0624, compared with -0.0623 for the pretrained encoder. The median tune errors are 0.08% and 0.03% over three seeds. The oscillator is therefore linearly observable with either pretrained or random encoder weights. A mean-pooled observable selects the second harmonic at large amplitude (Figure 3d), so a signed probe is required.

Under repeated prediction, the horizon is $\tau \approx 15$ to 21 steps under the cosine-similarity criterion in Appendix D (Figure 1d). The leading Lyapunov time is $1/\lambda_1 \approx 6.7$ steps. The injected frequency then enters the model's intrinsic band, and the frequency diffusion along a long predicted trajectory is $51\times$ that of a regular reference. Rolling context windows of two and four frames give the same loss of the signal observable in the encoder representations.

The same measurements are applied to the DeepMind Hamiltonian Dynamics Suite (Botev et al., 2021) to test transfer beyond the injected oscillator (Appendix B). Each 32×32 frame is bilinearly upsampled to 256^2 and passed through the frozen ViT-g backbone without fine-tuning. Frequency recovery is tested on 60 pendulum and 40 mass-spring trajectories. Chaos classification uses 50 pendulum and 50 double-pendulum trajectories, and dissipation tests use friction-0.05 pendulums with conservative controls. The frozen representations show per-frame motion $6.6\times$ above the random-pair floor and contain a component correlated with the true angle at $|r| = 0.82$.

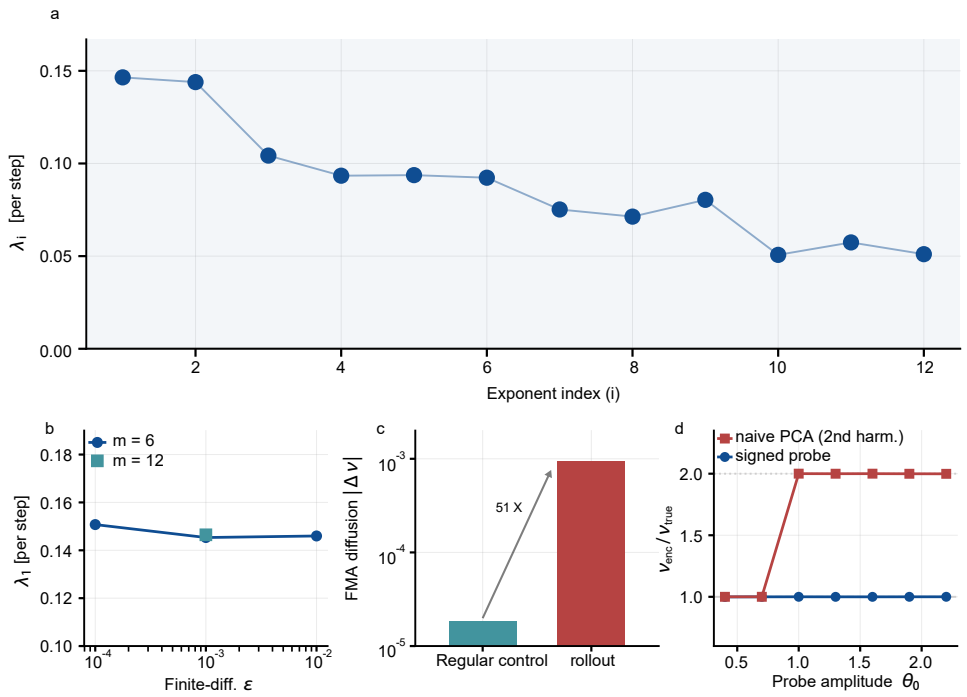


Figure 3: Observability in the encoder and instability in the predictor in V-JEPA 2 (frozen ViT-g encoder, 0.3B predictor iterated as an autonomous latent map). **(a)** All 12 estimated finite-time Lyapunov exponents are positive ($0.146 \rightarrow 0.051$). **(b)** λ_1 is stable across two decades of the finite-difference scale ϵ and for $m = 6$ and 12 tangent vectors. **(c)** FMA frequency diffusion of a long prediction is $51\times$ that of a regular control. **(d)** The signed probe tracks the fundamental ($\nu_{\text{enc}}/\nu_{\text{true}} = 1.000$) across amplitude, while a PCA observable selects the second harmonic (2.000) at large amplitude.

The pendulum dispersion is recovered with a median residual of 0.011%, while the mass-spring control remains amplitude independent (Figure A2a to c). Frequency diffusion separates regular pendulums from chaotic double pendulums by a median factor of 112 and gives an AUC of 0.98, compared with 0.96 from a single true angle (Figure A2d to f). For the friction pendulum, the aligned latent probe recovers the damping rate with correlation 0.9995 and classifies dissipation with AUC 1.00. The top principal component instead follows the randomized bob colour and gives AUC 0.55 (Figure A2g to i). The three diagnostics transfer across conservative, chaotic, and dissipative systems, provided that the observable is aligned with the dynamics.

Figure 4a shows the external-data test on a 661-frame laboratory-pendulum video spanning 18.7 periods. Two independent pixel trackers agree to 0.08° (Appendix H).

A linear probe fitted on the first 60% of the clip decodes the angle on the held-out remainder at $R^2 = 0.976$ (fit 0.998). The decoded frequency matches pixel tracking to 0.04%; across sliding windows, the median error is 0.05%. The frequency changes as the amplitude decays. The latent probe recovers the tune-shift coefficient at -0.060 ± 0.020, compared with the pixel value -0.059 ± 0.005 and the analytic $-1/16$ (Figure 4b). The top-component observable gives exactly $2.000\times$ the true frequency, reproducing the second-harmonic selection in Figure 3d on real data.

The pixel trajectory gives $\gamma_{\text{true}} = 3.84 \times 10^{-4}$ per frame per dimension, corresponding to $Q \approx 231$. A one-step latent fit gives 4.8×10^{-3}. The errors-in-variables bias predicted by the attenuation analysis therefore exceeds the physical rate by a factor of 12. Direct regressions of s_{t+n} on s_t apply this attenuation once, so its contribution decays as $1/n$ (Lemma A.17). The $n = 1$ excess of 4.40×10^{-3} corresponds to $\sigma = 6.6\%$, between independent estimates of 6.0% and approximately 7%. At $n = 420$, the latent estimate converges to $\gamma = 3.49 \times 10^{-4}$ with a 95% block-bootstrap

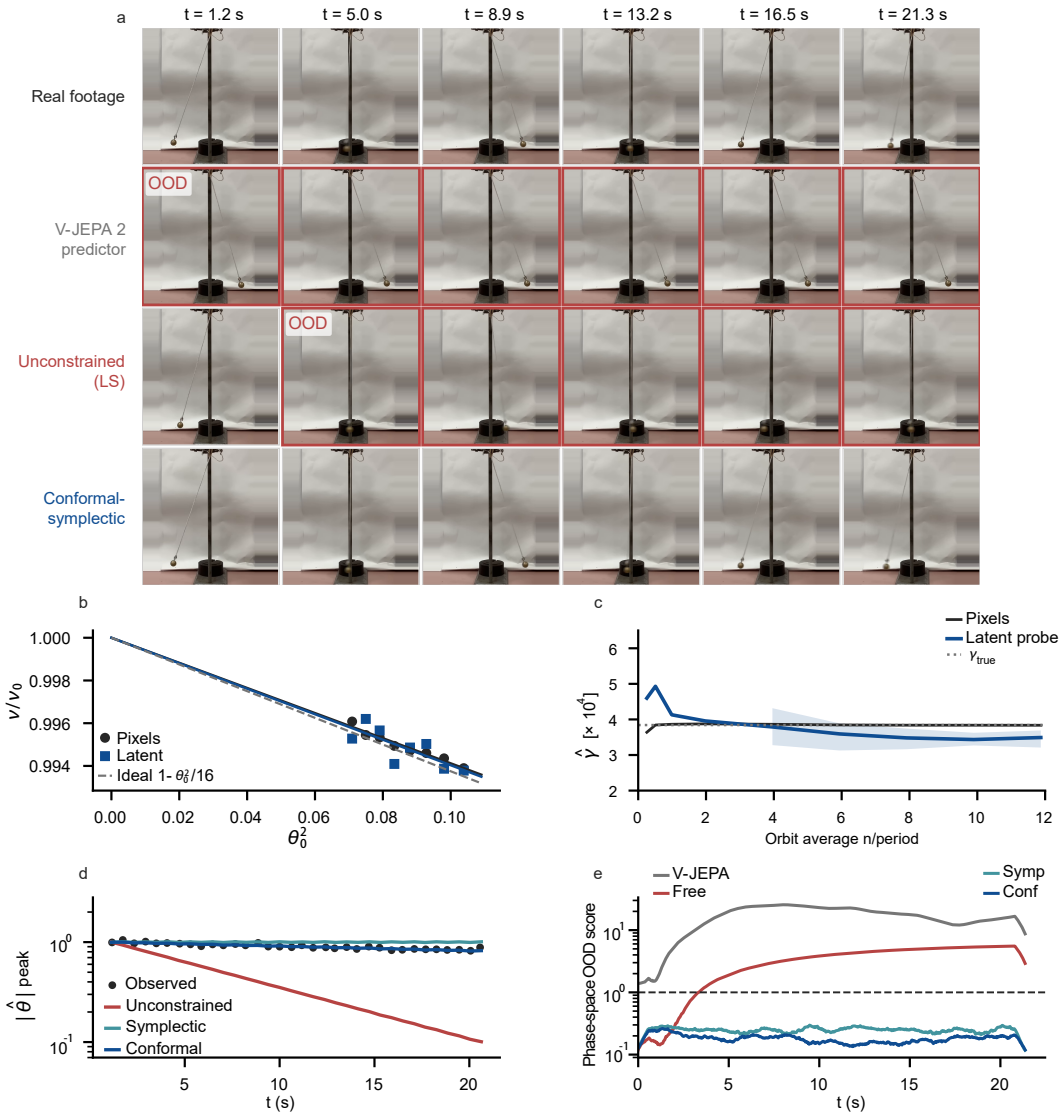


Figure 4: Autonomous prediction of a real pendulum for 18.4 periods in the representations of the frozen V-JEPA 2 encoder. **(a)** Four predicted trajectories rendered at six times. Each panel retrieves the observed frame nearest to the predicted angle, so only the dynamics differ among rows. A red border marks a phase state absent from the real trajectory. **(b)** The encoder recovers the $-1/16$ tune shift (latent -0.060 ± 0.020; pixel truth -0.059 ± 0.005). **(c)** The orbit-averaged contraction rate converges to $\gamma = 3.49 \times 10^{-4}$ (95% CI $[3.21, 3.69] \times 10^{-4}$), compared with the pixel value 3.84×10^{-4}. **(d)** Imposing the measured γ in the conformal-symplectic predictor reproduces the terminal amplitude (0.82 predicted and observed; unconstrained 0.10; symplectic 1.00). **(e)** V-JEPA 2's predictor and the unconstrained fit are off-manifold for 98% and 86% of frames, respectively; the de-biased and structure-preserving predictions remain on-manifold. TLS- and forward-backward DMD traces overlap the conformal trace at amplitude 0.85 and are omitted for clarity. Table 1 compares all eight predictors.

interval of $[3.21, 3.69] \times 10^{-4}$ (Figure 4c). This value is 9% below the pixel result, consistent with the 8.5% offset from a model-free envelope fit.

Table 1 compares autonomous predictors using the measured contraction rate. The V-JEPA 2 prediction seeded from the real video loses the pendulum after $\tau = 21$ steps, or 0.6 period. This falls within the range of 15 to 21 steps measured in the synthetic probes. The linear predictors are fitted

in a 2-D delay embedding with a quarter-period delay and condition number 1.03. The baselines are exact DMD, TLS-DMD (Hemati et al., 2017), forward-backward and noise-corrected DMD (Dawson et al., 2016), symplectic DMD, and conformal-symplectic DMD with the measured γ. The unconstrained eigenvalue magnitude is 0.996072. After removal of the physical factor $e^{-\gamma_{\mathrm{true}}}$, the residual attenuation corresponds to $\sigma = 6.0\%$, close to the approximately 7% estimate from the fit bias.

TLS- and forward-backward DMD remove the artificial damping and remain on the observed manifold. Their inferred dissipation is 25% below the pixel value, compared with a 9% offset for the conformal predictor. This residual includes finite-sample error, amplitude-dependent drift, and the probe's approximately 10% systematic error (Appendix H). MEND supplies the contraction rate from the coordinate-invariant long-lag measurement, while the structural parameterization fixes the reciprocal spectral class. The two noise-covariance corrections depend explicitly on σ^2 and give excessive growth or divergence for the available noise estimates. Only the V-JEPA 2 and unconstrained predictions leave the observed phase-space manifold in Figure 4d, e.

Table 1: Predictor comparison on the real pendulum over 651 autonomous prediction steps (18.4 periods). The linear maps are fitted in the same 2-D delay embedding of the probe observable. The implied contraction rate is $\gamma_{\mathrm{impl}} = -\log|\det M|/2$; the pixel reference is 3.84×10^{-4}. The terminal amplitude (amp) uses a fitted exponential envelope for every trajectory, including the observed reference. OOD is the fraction of steps whose phase state lies more than four median nearest-neighbour distances from any observed state. The fitted maps recover the frequency to 0.20% and differ mainly in amplitude decay. Appendix H describes the fitting procedure and estimators.

predictor	$\det M$	$\lvert\mathrm{eig}\rvert$	γ_{impl}	amp	OOD
observed (truth)	N/A	N/A	3.84×10^{-4}	0.82	N/A
V-JEPA 2's own predictor	N/A	N/A	N/A	N/A	98%
unconstrained LS = exact DMD	0.99216	0.99607	3.94×10^{-3}	0.10	86%
TLS-DMD (Hemati et al., 2017)	0.99942	0.99971	2.89×10^{-4}	0.85	0%
fb-DMD (Dawson et al., 2016)	0.99943	0.99971	2.88×10^{-4}	0.85	0%
nc-DMD, σ^2 from log-det	1.00099	1.00050	-4.96×10^{-4}	1.34	66%
nc-DMD, σ^2 from probe residual	1.04167	1.02062	-2.04×10^{-2}	diverges	100%
symplectic ($\det = 1$)	1.00000	1.00000	0	1.00	0%
MEND conformal-symplectic	0.99930	0.99965	3.49×10^{-4}	0.82	0%

The frozen encoder retains the oscillator observable, while predictor iteration determines how long that observable follows the encoded trajectory. The estimate $\tau \approx \lambda_1^{-1} \log(\Delta/\delta_0)$ depends primarily on the predictor exponent, with encoder errors entering logarithmically. The measured product $\tau\lambda_1 \approx 2$ to 3 gives horizons of 15 to 21 steps for Lyapunov times of 6 to 23 steps. Proposition A.3 and Theorem A.22 therefore motivate spectral control of open-loop growth. V-JEPA 2-AC uses receding-horizon planning and re-encodes at every step in its standard operation; the experiments here examine autonomous prediction. Extending elliptic or conformal structure from a 2×2 observable to a token-field predictor remains an open problem.

The signed-angle probe distinguishes the physical oscillation from harmonics and visual variation such as bob colour. Selecting this observable is therefore part of diagnosing the encoder and predictor.

Geometric and distributional constraints act on different parts of the model. SIGReg and VICReg preserve variation in the embedding distribution, whereas symplectic structure constrains the encoder map and predictor spectrum. On the unregularized T_0 system, SIGReg, VICReg, and the symplectic encoder all retain rank 4. Under observation noise, the unconstrained predictors paired with SIGReg or VICReg contract to $1/(1+\sigma^2)$, while the symplectic predictor remains on the unit circle. Combining SIGReg with a symplectic predictor preserves both full rank and unit-modulus eigenvalues.

The physical regime determines the predictor's spectral class. Conservative dynamics admit an elliptic symplectic predictor. Uniformly dissipative dynamics admit the conformal class after γ is measured from the orbit-averaged log-determinant. An isotropy constraint cannot supply this rate. On dissipative T_0 data, SIGReg reduces the latent isotropy score to 1.7 but gives 155% error when

γ is read through a predictor with a learned scale factor. The free-predictor log-determinant gives 2.4% error. Anisotropic dissipation requires per-mode contraction factors, since one conformal factor cannot represent unequal decay rates. The observable must contain the corresponding phase-space variables for its spectrum to describe their dynamics.

The theory treats linear symplectic dynamics, and the V-JEPA 2 experiments examine one frozen model under autonomous iteration. The real-video experiment identifies 2×2 predictors from one trajectory. Its latent contraction rate has a 9% systematic probe error, and its terminal amplitude varies by $\pm 8\%$ across estimators. Over 18 periods at $Q \approx 231$, the envelope does not distinguish among decay laws, so the constant-γ conformal model gives a first-order description. Contact-rich dynamics require anisotropic dissipative structure, and each target system requires independent FMA calibration.

4 Conclusion

An iterated latent predictor is a discrete dynamical system whose spectrum shapes long-horizon behavior. In the linear-symplectic setting, the semi-conjugacy formulation relates dimensional collapse to encoder rank and prediction drift to predictor spectrum. These two degenerate modes therefore require separate constraints. Under the tested V-JEPA 2 protocol, at least 12 positive finite-time Lyapunov exponents accompany an open-loop horizon of 15 to 21 steps. The oscillator remains observable in the representations of the frozen encoder. On one real video, MEND uses the measured contraction rate to constrain a conformal-symplectic predictor. It increases the open-loop horizon by a factor of 31 while following the observed decay in a 2-D pendulum observable. The diagnostics identify whether prediction failure originates in the representation or predictor and select a spectral constraint for low-dimensional observables. Extending these constraints to token-field predictors may support autonomous video prediction for planning, control, and scientific simulation.

Ethics Statement.

This study examines the dynamics of learned latent world models. The experiments use synthetic dynamical systems, a public physics benchmark (Botev et al., 2021), and a publicly released frozen model checkpoint (Assran et al., 2025). They also use a public pendulum recording from `teachphysics.org`, licensed under CC BY-NC-SA 4.0 and cropped with the creator's permission. No identifiable person appears in the frames used, and the study introduces no new data involving human subjects or personally identifiable information. We foresee no ethical concerns or harmful applications specific to this analysis beyond those general to representation and world-model research.

Reproducibility Statement.

Section 2 states the theoretical assumptions, and Appendix A gives the proofs. Appendix D specifies the systems, training procedures, and frozen V-JEPA 2 evaluation setup. Appendix E defines and calibrates the three dynamical measurements. The controlled T_0 studies use three fixed seeds and report medians. The V-JEPA 2 experiments use the public frozen V-JEPA 2-AC checkpoint with fixed action and proprioceptive conditioning. Appendix H describes the tracking, time units, estimator calibration, and bootstrap intervals for the real-video experiment. The code and data will be available via GitHub soon.

Generative AI disclosure

Generative AI tools assisted with translation, language editing, formatting references, LaTeX, and software code. They were not used to generate synthetic datasets, formulate mathematical claims, supply ingredients for proofs, or write proofs. They were not used to propose hypotheses, design the methodology or experiments, clean or analyze data, or interpret results. The other activities requiring disclosure do not apply to this work. The authors reviewed every AI-assisted edit and checked the code by inspection and testing. The authors take responsibility for the final text, claims, code, and other materials produced with AI assistance.

REFERENCES

Mahmoud Assran, Quentin Duval, Ishan Misra, Piotr Bojanowski, Pascal Vincent, Michael Rabbat, Yann LeCun, and Nicolas Ballas. Self-supervised learning from images with a joint-embedding predictive architecture. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15619–15629. IEEE, 2023.

Mido Assran, Adrien Bardes, David Fan, Quentin Garrido, Russell Howes, Matthew Muckley, Ammar Rizvi, Claire Roberts, Koustuv Sinha, Artem Zholus, et al. V-jepa 2: Self-supervised video models enable understanding, prediction and planning. *arXiv preprint arXiv:2506.09985*, 2025.

Omri Azencot, N Benjamin Erichson, Vanessa Lin, and Michael Mahoney. Forecasting sequential data using consistent koopman autoencoders. In *International Conference on Machine Learning (ICML)*, pp. 475–485. PMLR, 2020.

Randall Balestriero and Yann LeCun. Lejepa: Provable and scalable self-supervised learning without the heuristics. *arXiv preprint arXiv:2511.08544*, 2025.

Adrien Bardes, Jean Ponce, and Yann LeCun. VICReg: Variance-invariance-covariance regularization for self-supervised learning. *International Conference on Learning Representations (ICLR)*, 2022.

Adrien Bardes, Quentin Garrido, Jean Ponce, Xinlei Chen, Michael Rabbat, Yann LeCun, Mido Assran, and Nicolas Ballas. V-JEPA: Latent video prediction for visual representation learning. *arXiv preprint*, 2024.

Aleksandar Botev, Andrew Jaegle, Peter Wirnsberger, Daniel Hennes, and Irina Higgins. Which priors matter? Benchmarking models for learning latent dynamics. In *Advances in Neural Information Processing Systems (NeurIPS), Datasets and Benchmarks Track*, 2021. arXiv:2111.05458.

Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9630–9640. IEEE, 2021.

Boyuan Chen, Diego Martí Monsó, Yilun Du, Max Simchowitz, Russ Tedrake, and Vincent Sitzmann. Diffusion forcing: Next-token prediction meets full-sequence diffusion. *Advances in Neural Information Processing Systems (NeurIPS)*, 37:24081–24125, 2024.

Renyi Chen and Molei Tao. Data-driven prediction of general hamiltonian dynamics via learning exactly-symplectic maps. In *International Conference on Machine Learning (ICML)*, pp. 1717–1727. PMLR, 2021.

Xinlei Chen and Kaiming He. Exploring simple siamese representation learning. In *2021 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp. 15745–15753. IEEE, 2021.

Yongsheng Chen, Wei Guo, Qi Tang, and Xinghui Zhong. Reduced-order modeling of hamiltonian dynamics based on symplectic neural networks. *arXiv preprint arXiv:2508.11911*, 2025.

Scott TM Dawson, Maziar S Hemati, Matthew O Williams, and Clarence W Rowley. Characterizing and correcting for the effect of sensor noise in the dynamic mode decomposition. *Experiments in Fluids*, 57(3):42, 2016.

Luca Falorsi, Pim De Haan, Tim R Davidson, Nicola De Cao, Maurice Weiler, Patrick Forré, and Taco S Cohen. Explorations in homeomorphic variational auto-encoding. *arXiv preprint arXiv:1807.04689*, 2018.

Marc Finzi, Ke Alexander Wang, and Andrew G Wilson. Simplifying hamiltonian and lagrangian neural networks via explicit constraints. *Advances in neural information processing systems (NeurIPS)*, 33:13880–13889, 2020.

Quentin Garrido, Nicolas Ballas, Mahmoud Assran, Adrien Bardes, Laurent Najman, Michael Rabbat, Emmanuel Dupoux, and Yann LeCun. Intuitive physics understanding emerges from self-supervised pretraining on natural videos. *arXiv preprint arXiv:2502.11831*, 2025.

Samuel Greydanus, Misko Dzamba, and Jason Yosinski. Hamiltonian neural networks. *Advances in Neural Information Processing Systems (NeurIPS)*, 32, 2019.

Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Guo, Mohammad Gheshlaghi Azar, et al. Bootstrap your own latent-a new approach to self-supervised learning. *Advances in neural information processing systems (NeurIPS)*, 33:21271–21284, 2020.

Danijar Hafner, Timothy Lillicrap, Ian Fischer, Ruben Villegas, David Ha, Honglak Lee, and James Davidson. Learning latent dynamics for planning from pixels. In *International Conference on Machine Learning (ICML)*, pp. 2555–2565. PMLR, 2019.

Danijar Hafner, Timothy Lillicrap, Jimmy Ba, and Mohammad Norouzi. Dream to control: Learning behaviors by latent imagination. In *International Conference on Learning Representations (ICLR)*, 2020.

Danijar Hafner, Jurgis Pasukonis, Jimmy Ba, and Timothy Lillicrap. Mastering diverse control tasks through world models. *Nature*, 640(8059):647–653, 2025.

Nick Hansen, Hao Su, and Xiaolong Wang. Td-mpc2: Scalable, robust world models for continuous control. In *International Conference on Learning Representations (ICLR)*, volume 2024, pp. 47376–47405, 2024.

Maziar S Hemati, Clarence W Rowley, Eric A Deem, and Louis N Cattafesta. De-biasing the dynamic mode decomposition for applied koopman spectral analysis of noisy datasets. *Theoretical and Computational Fluid Dynamics*, 31(4):349–368, 2017.

Irina Higgins, Andrew Jaegle, Peter Wirnsberger, and Aleksandar Botev. SyMetric: Measuring the quality of learnt Hamiltonian dynamics inferred from vision. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021. arXiv:2111.05986.

Pengzhan Jin, Zhen Zhang, Aiqing Zhu, Yifa Tang, and George Em Karniadakis. SympNets: Intrinsic structure-preserving symplectic networks for identifying hamiltonian systems. *Neural Networks*, 2020.

David Klindt, Yann LeCun, and Randall Balestriero. When does lejepa learn a world model? *arXiv preprint arXiv:2605.26379*, 2026.

Anurag Koul, Shivakanth Sujit, Shaoru Chen, Ben Evans, Lili Wu, Byron Xu, Rajan Chari, Riashat Islam, Raihan Seraj, Yonathan Efroni, Lekan P. Molu, Miroslav Dudík, John Langford, and Alex Lamb. PcLast: Discovering plannable continuous latent states. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 25475–25493, 2024. URL `https://proceedings.mlr.press/v235/koul24a.html`.

Jacques Laskar. Frequency map analysis and quasiperiodic decompositions. *Hamiltonian Systems and Fourier Analysis*, 1993.

Yann LeCun et al. A path towards autonomous machine intelligence version 0.9. 2, 2022-06-27. *Open Review*, 62(1):1–62, 2022.

Xinyi Li, Zaishuo Xia, Chenjie Hao, and Yubei Chen. Beyond myopic world models: Long-horizon end-to-end training for direct future prediction. *arXiv preprint arXiv:2608.07420*, 2026.

Hongyuan Adam Lu, ZL Wei, Qun Zhang, Jinrui Zeng, Bowen Cao, Lingwei Meng, Mocheng Li, Zezhong Wang, Haonan Yin, Naifu Xue, et al. Looped world models. *arXiv preprint arXiv:2606.18208*, 2026.

Bethany Lusch, J Nathan Kutz, and Steven L Brunton. Deep learning for universal linear embeddings of nonlinear dynamics. *Nature Communications*, 9(1):4950, 2018.

Lucas Maes, Quentin Le Lidec, Damien Scieur, Yann LeCun, and Randall Balestriero. Leworld-model: Stable end-to-end joint-embedding predictive architecture from pixels. *arXiv preprint arXiv:2603.19312*, 2026.

Samuel E Otto and Clarence W Rowley. Linearly recurrent autoencoder networks for learning dynamics. *SIAM Journal on Applied Dynamical Systems*, 18(1):558–593, 2019.

Zhongwei Ren, Yunchao Wei, Xun Guo, Yao Zhao, Bingyi Kang, Jiashi Feng, and Xiaojie Jin. Videoworld: Exploring knowledge learning from unlabeled videos. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 29029–29039. IEEE, 2025.

Julian Schrittwieser, Ioannis Antonoglou, Thomas Hubert, Karen Simonyan, Laurent Sifre, Simon Schmitt, Arthur Guez, Edward Lockhart, Demis Hassabis, Thore Graepel, et al. Mastering atari, go, chess and shogi by planning with a learned model. *Nature*, 588(7839):604–609, 2020.

Naoya Takeishi, Yoshinobu Kawahara, and Takehisa Yairi. Learning Koopman invariant subspaces for dynamic mode decomposition. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

Peter Toth, Danilo Jimenez Rezende, Andrew Jaegle, Sébastien Racanière, Aleksandar Botev, and Irina Higgins. Hamiltonian generative networks. *arXiv preprint arXiv:1909.13789*, 2019.

Jialong Wu, Shaofeng Yin, Ningya Feng, Xu He, Dong Li, Jianye Hao, and Mingsheng Long. ivideogpt: Interactive videogpts are scalable world models. *Advances in Neural Information Processing Systems (NeurIPS)*, 37:68082–68119, 2024.

Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, and Stéphane Deny. Barlow twins: Self-supervised learning via redundancy reduction. In *International Conference on Machine Learning (ICML)*, pp. 12310–12320. PMLR, 2021.

Conglang Zhang, Yifan Zhan, Qingjie Wang, Zhanpeng Ouyang, Yu Li, Zihao Yang, Xiaoyang Guo, Weiqiang Ren, Qian Zhang, Zhen Dong, et al. Horizondrive: Self-corrective autoregressive world model for long-horizon driving simulation. *arXiv preprint arXiv:2605.11596*, 2026.

Kai Zhao, Dongliang Nie, Yuchen Lin, Zhehan Luo, Yixiao Gu, Deng-Ping Fan, and Dan Zeng. Sub-jepa: Subspace gaussian regularization for stable end-to-end world models. *arXiv preprint arXiv:2605.09241*, 2026.

Yaofeng Desmond Zhong and Naomi Leonard. Unsupervised learning of lagrangian dynamics from images for prediction and control. *Advances in neural information processing systems (NeurIPS)*, 33:10741–10752, 2020.

Yaofeng Desmond Zhong, Biswadip Dey, and Amit Chakraborty. Dissipative symoden: Encoding hamiltonian dynamics with dissipation and control into deep learning. *arXiv preprint arXiv:2002.08860*, 2020.

Gaoyue Zhou, Hengkai Pan, Yann Lecun, and Lerrel Pinto. DINO-WM: World models on pre-trained visual features enable zero-shot planning. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning (ICML)*, volume 267 of *Proceedings of Machine Learning Research*, pp. 79115–79135. PMLR, 13–19 Jul 2025. URL https://proceedings.mlr.press/v267/zhou25t.html.

A FORMAL STATEMENTS AND PROOFS

Remark A.1: scope of the semi-conjugacy model. Four features of JEPA training are compressed into (E, P, Φ). First, $\bar{E}$ is an exponential moving average of E, so that the defect δ includes the encoder lag and vanishes only at an EMA fixed point. Second, masked-target training averages over mask patterns, whereas the autonomous prediction uses the fully observed member of that family. Third, fixing (a^*, s^*) selects one autonomous map from the action-conditioned family $P(z, a, s)$. Finally, the native input is a multi-frame clip; the tests use context lengths $k = 1, 2, 4$ (Appendix D). These distinctions delimit the application of the linear-regime theorems to the trained model.

This section gives the symplectic-subspace preliminaries, induced-form definition, formal statements, and proofs for the four structural implications summarized in §2. Theorem and equation numbering continues the body sequence.

Assumption A.2: linear theory. (i) The true dynamics is linear, deterministic and symplectic: $\Phi(x) = Rx$ with $R \in \mathrm{Sp}(2d)$ invertible. (ii) The encoder is linear, $E(x) = Cx$ with $C \in \mathbb{R}^{m\times 2d}$, $1 \le m \le 2d$; unless stated otherwise C has full row rank m. (iii) The predictor is linear, $P(z) = Mz$ with $M \in \mathbb{R}^{m\times m}$. (iv) Semi-conjugacy means $CR = MC$, the population optimum of the JEPA objective $\|P(E(x)) - E(\Phi(x))\|^2$ in the linear class.

Proposition A.3: prediction error amplification. Suppose $\delta(x) \le \delta_0$ along an orbit $x, \Phi(x), \ldots, \Phi^{\circ n-1}(x)$, and let P be L-Lipschitz on a region containing both the predicted orbit $\{P^{\circ j}(E(x))\}_{j=0}^{n-1}$ and the encoded true orbit $\{E(\Phi^{\circ j}(x))\}_{j=0}^{n-1}$. Then the prediction error satisfies (3).

Proof of Proposition A.3. Let $e_k := \big\|P^{\circ k}(E(x)) - E(\Phi^{\circ k}(x))\big\|$, so $e_0 = 0$. By the triangle inequality,

$$\begin{aligned} e_{k+1} &\le \big\|P\big(P^{\circ k}(E(x))\big) - P\big(E(\Phi^{\circ k}(x))\big)\big\| \\ &\quad + \big\|P\big(E(\Phi^{\circ k}(x))\big) - E(\Phi^{\circ k+1}(x))\big\| \\ &\le L\, e_k + \delta(\Phi^{\circ k}(x)) \le L\, e_k + \delta_0, \end{aligned}$$

where the Lipschitz bound is applied to the pair of points $P^{\circ k}(E(x))$ and $E(\Phi^{\circ k}(x))$, both in the assumed region. Unrolling the recursion gives $e_n \le \delta_0 \sum_{j=0}^{n-1} L^j$.

Preliminaries. We work on the phase space $(\mathbb{R}^{2d}, \Omega)$ with the standard symplectic matrix $\Omega = \left(\begin{smallmatrix} 0 & I_d \\ -I_d & 0 \end{smallmatrix}\right)$ and symplectic group $\mathrm{Sp}(2d) = \{R \in \mathbb{R}^{2d\times 2d} : R^\top \Omega R = \Omega\}$. For a subspace $W \subseteq \mathbb{R}^{2d}$, its symplectic orthogonal is $W^\Omega = \{u : \Omega(u, w) = 0 \text{ for all } w \in W\}$; W is symplectic if $\Omega|_W$ is nondegenerate, equivalently $W \cap W^\Omega = 0$, equivalently $\mathbb{R}^{2d} = W \oplus W^\Omega$.

Definition A.4: induced latent form. Let $W = \ker C$ be R-invariant. The encoder induces a bilinear form on the latent space from Ω if the pullback condition

$$\omega_Z(Cu,\, Cv) \;=\; \Omega(u, v) \qquad \text{for all } u, v \in W^\Omega \tag{A1}$$

determines a (necessarily unique) form ω_Z on all of $\mathbb{R}^m$; we write Ω_Z for its matrix.

The radical of $\Omega|_{W^\Omega}$ is $W \cap W^\Omega$, which equals the kernel of $C|_{W^\Omega}$. The form therefore descends to the quotient and defines a nondegenerate form on $C(W^\Omega)$. It uniquely determines a form on the entire latent space exactly when W is symplectic, or equivalently when $W \cap W^\Omega = 0$ and $C(W^\Omega) = \mathbb{R}^m$.

Lossless latent reduction: admissible encoders and inseparable canonical pairs.

Proposition A.5 (existence and uniqueness of the latent propagator). Under Assumption A.2, there exists $M \in \mathbb{R}^{m\times m}$ with $CR = MC$ if and only if $\ker C$ is R-invariant, $R(\ker C) = \ker C$. In that case M is unique and equals $M = CRC^+$ for any right inverse C^+ of C.

Proof. The equation $MC = CR$ is solvable in M iff $\ker C \subseteq \ker(CR)$ (a linear map factors through C iff it kills $\ker C$). Since $\ker(CR) = R^{-1}(\ker C)$, the condition reads $R(\ker C) \subseteq \ker C$, and equality follows from invertibility of R and finite dimension. For uniqueness, C is surjective, so M is determined by its values on $\operatorname{im} C = \mathbb{R}^m$; the formula follows from $MCC^+ = CRC^+$ and $CC^+ = I_m$. □

Theorem A.6: lossless symplectic reduction. Under Assumption A.2 with $W = \ker C$ R-invariant, the following are equivalent: (i) W is a symplectic subspace; (ii) the latent inherits a form ω_Z via (A1), and this ω_Z is nondegenerate. When these hold, the induced propagator M of Proposition A.5 is ω_Z-symplectic, $M^\top \Omega_Z M = \Omega_Z$.

Proof. $(i) \Rightarrow (ii)$. If W is symplectic then $\mathbb{R}^{2d} = W \oplus W^\Omega$ and $\Omega|_{W^\Omega}$ is nondegenerate. Since $\ker C = W$, the restriction $C|_{W^\Omega} : W^\Omega \to \mathbb{R}^m$ is injective, and by dimension count ($\dim W^\Omega = 2d - \dim W = m$) an isomorphism; (A1) transports $\Omega|_{W^\Omega}$ to a well-defined nondegenerate form on $\mathbb{R}^m$. $(ii) \Rightarrow (i)$. In general $\dim C(W^\Omega) = \dim W^\Omega - \dim(W^\Omega \cap \ker C) = m - \dim(W \cap W^\Omega)$. Condition (A1) constrains ω_Z only on $C(W^\Omega)$; it determines ω_Z on all of $\mathbb{R}^m$ iff $C(W^\Omega) = \mathbb{R}^m$, i.e. iff $W \cap W^\Omega = 0$, i.e. iff W is symplectic. (If $W \cap W^\Omega \neq 0$, any extension of the transported form is an arbitrary gauge choice.) Symplecticity of M: $R \in \mathrm{Sp}(2d)$ preserves Ω-orthogonality, so $RW = W$ implies $RW^\Omega = W^\Omega$; $R|_{W^\Omega}$ preserves $\Omega|_{W^\Omega}$, and M is its conjugate under the isomorphism $C|_{W^\Omega}$, hence preserves ω_Z. □

Remark A.7. A coisotropic kernel ($W^\Omega \subseteq W$) admits the classical linear symplectic reduction W/W^Ω; the symplectic-kernel case above is the one in which the reduction is realized by the encoder itself, with no quotient.

Theorem A.8: canonical pairs are inseparable. Let $R = \bigoplus_{j=1}^{d} \mathrm{Rot}(\mu_j)$ act on the canonical planes $\Pi_j = \mathrm{span}(q_j, p_j)$ in normalized coordinates, with tunes $\mu_j \in (0, \pi)$ pairwise distinct. Then: (a) every R-invariant subspace of $\mathbb{R}^{2d}$ is a direct sum $\bigoplus_{j \in S} \Pi_j$ for some $S \subseteq \{1, \ldots, d\}$; (b) consequently a lossless symplectic reduction (Theorem A.6) exists iff $\ker C = \bigoplus_{j \in S} \Pi_j$: whole canonical planes are retained or discarded, never split; (c) the minimal latent dimension that losslessly carries k retained degrees of freedom is exactly $2k$; in particular no odd-dimensional latent, and no "position without momentum," admits a lossless symplectic reduction.

Proof. (a) Complexify. The eigenvalues $e^{\pm i\mu_j}$ are $2d$ distinct numbers, so every complex R-invariant subspace is a sum of one-dimensional eigenspaces; real invariant subspaces correspond to conjugation-closed sets of eigenvalues, and $\{e^{i\mu_j}, e^{-i\mu_j}\}$ spans exactly $\Pi_j \otimes \mathbb{C}$. (b) Each Π_j is symplectic and the Π_j are mutually Ω-orthogonal, so $\bigoplus_{j \in S} \Pi_j$ is symplectic; by (a) there are no other invariant candidates, and any isotropic direction (a single q_j-line) is excluded because an odd-dimensional subspace, or a proper line of a plane, is never symplectic. (c) follows from (b) by dimension count. □

Remark A.9: degenerate tunes. If tunes coincide ($\mu_i = \mu_j$) the invariant-subspace lattice is larger (any plane inside the corresponding isoclinic block is invariant) and mixed reductions become possible; the parity statement (c) and the inseparability of a nondegenerate pair survive. Distinctness of tunes is generic.

Encoder-side nondegeneracy: why a symplectic encoder prevents collapse.

Theorem A.10 (nondegeneracy excludes dimensional collapse). Drop the full-rank assumption on C. If the latent inherits a nondegenerate ω_Z in the sense of Definition A.4, then $\mathrm{rank} C = m$. Equivalently, the pullback condition (A1) determines a nondegenerate ω_Z on $\mathbb{R}^m$ if and only if $\mathrm{rank} C = m$ and $\ker C$ is symplectic.

Proof. Let $r = \mathrm{rank} C$ and $W = \ker C$, $\dim W = 2d - r$, so $\dim W^\Omega = r$ and a dimension count gives $\dim C(W^\Omega) = r - \dim(W \cap W^\Omega) \leq r$. If $r < m$ the pullback condition constrains ω_Z on a subspace of dimension $\leq r < m$ and cannot determine, let alone render nondegenerate, a form

on $\mathbb{R}^m$; hence $r = m$. Given $r = m$, nondegeneracy is equivalent to $\ker C$ being symplectic by Theorem A.6. A rank-deficient representation therefore cannot induce a nondegenerate symplectic form on all of $\mathbb{R}^m$. □

Remark A.11: limits of the full-rank condition. Theorem A.10 excludes dimensional collapse but not poor conditioning, folding, or loss of semantics. The symplectic encoder has condition number 1.08, compared with approximately 10^{29} for the collapsed free encoder and 55 when the encoder absorbs dissipation into its gauge. The bounded-distortion condition of Theorem A.24 supplies the corresponding quantitative control.

Corollary A.12: architectural anti-collapse. An encoder restricted to the symplectic-submersion class satisfies $\mathrm{rank} C = m$ throughout training. No stop-gradient, teacher-student asymmetry, or distributional regularizer is required to exclude collapse.

Remark A.13. A symplectic-submersion encoder remains full rank throughout training. In Exp. 1, the unconstrained encoder loses rank with either predictor.

Conjecture A.14: nonlinear extension. Let E be smooth with $W_x := \ker \mathrm{D}E(x)$ of constant rank, and suppose that at each x in an open set of data the pullback condition $\omega_Z\big(\mathrm{D}E(x)\,u,\, \mathrm{D}E(x)\,v\big) = \Omega(u, v)$ for all $u, v \in W_x^\Omega$ determines a nondegenerate two-form ω_Z on the latent. Then $\mathrm{D}E(x)$ has full rank m and W_x is a symplectic subspace; conversely, rank deficiency of $\mathrm{D}E(x)$ prevents the pullback condition from uniquely determining a nondegenerate latent two-form (any nondegenerate extension beyond the image of W_x^Ω is an arbitrary gauge choice, as in Theorem A.10). The forward direction is Theorem A.10 applied to $\mathrm{D}E(x)$; a global statement under training dynamics is open.

Predictor attenuation and stability.

Theorem A.15: errors-in-variables shrinkage. Let $z_{t+1} = Rz_t$, where $R \in \mathrm{Sp}(2d)$ and $RR^\top = I$, so $R \in \mathrm{Sp}(2d) \cap O(2d)$. Assume the state has stationary isotropic covariance $\mathbb{E}[z_t z_t^\top] = sI$, with $s > 0$. The orthogonality of R gives $R(sI)R^\top = sI$. The observations are $x_t = z_t + \varepsilon_t$, where $\varepsilon_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2 I)$ is independent of the states and across time. Draw N i.i.d. state pairs (z, Rz) from the stationary ensemble, the distribution $\mathcal{D}$ of §2. Then least-squares regression of x_{t+1} on x_t converges in probability as $N \to \infty$ to the population map $\widehat{M}_\infty$ in (4). Every eigenvalue of the estimate is the true eigenvalue shrunk radially by $(1 + \sigma^2/s)^{-1} < 1$: an n-step prediction damps amplitudes by $(1 + \sigma^2/s)^{-n}$ although the true dynamics conserves them. Throughout the empirical sections we take $s = 1$, so $1/(1 + \sigma^2/s) = 1/(1 + \sigma^2)$ as reported in the abstract and Table A2.

Proof. Independence of the noise across time and from the state gives $\mathbb{E}[x_{t+1} x_t^\top] = R\,\mathbb{E}[z_t z_t^\top] = sR$ and $\mathbb{E}[x_t x_t^\top] = (s + \sigma^2)I$; the ratio is (4), and convergence of the empirical moments as $N \to \infty$ is the law of large numbers for i.i.d. pairs. Radial shrinkage is immediate since $\widehat{M}_\infty$ is a scalar multiple of R. □

Remark A.16. A stationary state covariance satisfies $R\Sigma R^\top = \Sigma$. For the isotropic covariance $\Sigma = sI$ in Theorem A.15, this requires $R \in O(2d)$. Thus $R \in \mathrm{Sp}(2d) \cap O(2d)$, with all $|\lambda(R)| = 1$; the T_0 system belongs to this rotation class. A hyperbolic symplectic map cannot preserve a positive-definite state covariance. For an anisotropic covariance $\Sigma = \mathbb{E}[zz^\top]$ satisfying $R\Sigma R^\top = \Sigma$, the regression limit is $R\Sigma(\Sigma + \sigma^2 I)^{-1}$. The attenuation then depends on direction.

The contraction readout introduced under coordinate-identifiable dissipation is estimated by direct n-step regression. The following lemma gives its finite-lag statistical bias; coordinate invariance is treated separately by Theorem A.24.

Lemma A.17: errors-in-variables consistency of the n-step regression. Fix a window $t = 0, \ldots, L - 1$ and suppose (i) the latent state obeys $z_{t+1} = Az_t$ with $A \in \mathbb{R}^{m \times m}$ invertible and time-invariant on the window; (ii) observations are $x_t = z_t + \varepsilon_t$ with ε_t i.i.d., mean zero, covariance $\sigma^2 I$, independent of the states and across time; (iii) the window-pooled state second moment $\bar{\Sigma} :=$

$\frac{1}{L}\sum_t \mathbb{E}[z_t z_t^\top]$ is finite and positive definite. Let $\widehat{M}_n$ be the least-squares regression of x_{t+n} on x_t pooled over the window, $n \geq 1$. Then in the population limit

$$\widehat{M}_{n,\infty} \;=\; A^n\,\bar{\Sigma}\,(\bar{\Sigma}+\sigma^2 I)^{-1}, \tag{A2}$$

so that the per-step, per-dimension contraction readout satisfies

$$\widehat{\gamma}_n \;:=\; -\frac{1}{n\,m}\log\bigl|\det \widehat{M}_{n,\infty}\bigr| \;=\; \gamma \;+\; \frac{\beta}{n\,m}, \qquad \beta := \log\det\bigl(I+\sigma^2\bar{\Sigma}^{-1}\bigr) \;\geq\; 0, \tag{A3}$$

where $\gamma = -\frac{1}{m}\log|\det A|$ is the true contraction rate. The bias is nonnegative (the readout overestimates dissipation), its numerator β does not depend on n, and it therefore decays as $O(1/n)$. In the isotropic case $\bar{\Sigma} = sI$ it is exactly $\frac{1}{n}\log(1+\sigma^2/s)$.

Proof. Independence of the noise from the state and across time gives, for each t and each $n \geq 1$, $\mathbb{E}[x_{t+n}x_t^\top] = \mathbb{E}[z_{t+n}z_t^\top] = A^n\,\mathbb{E}[z_t z_t^\top]$ and $\mathbb{E}[x_t x_t^\top] = \mathbb{E}[z_t z_t^\top] + \sigma^2 I$. Averaging each over the window, the cross-moment pools to $A^n\bar{\Sigma}$ and the auto-moment to $\bar{\Sigma}+\sigma^2 I$; the pooled least-squares limit is their ratio, which is (A2). Taking $\log|\det(\cdot)|$ and using multiplicativity, $\log\left|\det \widehat{M}_{n,\infty}\right| = n\log|\det A| + \log\det[\bar{\Sigma}(\bar{\Sigma}+\sigma^2 I)^{-1}] = -nm\gamma - \beta$, which is (A3). □

Remark A.18: direct fit versus iterated fit. Iterating the one-step fit, $\widehat{P}^{\,n}$ with $\widehat{P} = \widehat{M}_1$, gives $\log\left|\det \widehat{P}^{\,n}\right| = n\log\left|\det \widehat{M}_1\right|$ and the per-step readout $\gamma + \beta/m$ for every n. By contrast, direct n-step regression gives the $1/n$ decay in (A3), as observed in §3.

Remark A.19: assumptions of Lemma A.17. Stationarity is unnecessary because both pooled moments use the same $\bar{\Sigma}$, so (A2) also holds for a decaying covariance sequence. Time invariance of A remains approximate: over 420 frames the pendulum amplitude falls from 0.35 to 0.25 rad, giving a 0.4% frequency drift from the $-1/16$ law. On the pixel-tracked angle, the resulting $\widehat{\gamma}_n$ remains within 0.9% of the truth for lags of 18 to 210 frames (Appendix H). For the delay embedding $s_t = (u_t, u_{t+\tau})$, the lag $n = \tau$ shares a noisy coordinate between regressor and target and is excluded from the trend.

Remark A.20: statistical and geometric conditions. Lemma A.17 gives $1/n$ convergence of the log-determinant estimate, while Theorem A.24 gives invariance under bounded-distortion changes of latent chart. Both conditions are required for a coordinate-independent estimate of γ.

Lemma A.21: reciprocal spectrum. If $M^\top \Omega_Z M = \Omega_Z$ with Ω_Z invertible, then $M^{-1} = \Omega_Z^{-1} M^\top \Omega_Z$; hence M^{-1} is similar to $M^\top$ and the spectrum of M is closed under $\lambda \mapsto 1/\lambda$ and (M being real) conjugation: eigenvalues occur in quadruples $\{\lambda, \bar{\lambda}, 1/\lambda, 1/\bar{\lambda}\}$.

Proof. From $M^\top \Omega_Z M = \Omega_Z$ we get $M^\top = \Omega_Z M^{-1}\Omega_Z^{-1}$, giving the stated similarity; $M^\top$ has the same spectrum as M, and similarity to M^{-1} closes it under $\lambda \mapsto 1/\lambda$. □

Theorem A.22: stability of iterated linear maps. For $M \in \mathbb{R}^{m\times m}$, $\sup_n \|M^n\| < \infty$ iff the spectral radius $\rho(M) \leq 1$ and every unit-modulus eigenvalue is semisimple. If in addition M is ω_Z-symplectic, then: (a) $|\det M| = 1$ (volume is preserved), yet boundedness may fail, since volume preservation is not stability; (b) $\sup_n \|M^n\| < \infty$ iff all eigenvalues lie on the unit circle and are semisimple (the elliptic case); (c) if some $|\lambda| < 1$, its reciprocal partner satisfies $|1/\lambda| > 1$ and the iteration diverges. A symplectic map therefore cannot produce uniform contraction without an expanding reciprocal mode.

Proof. The general criterion is the Jordan-form estimate $\|M^n\| \asymp n^{k-1}\rho(M)^n$ with k the largest Jordan block on the peripheral spectrum. (a) $\det M = \pm 1$ from Lemma A.21 (pairing $\lambda\cdot(1/\lambda)$), and a hyperbolic pair $(\lambda, 1/\lambda)$ with $|\lambda| \neq 1$ preserves volume while $\|M^n\| \to \infty$. (b) By Lemma A.21 the condition $\rho(M) \leq 1$ already forces all $|\lambda| = 1$; combined with the general criterion this gives the claim. (c) is Lemma A.21 read contrapositively. □

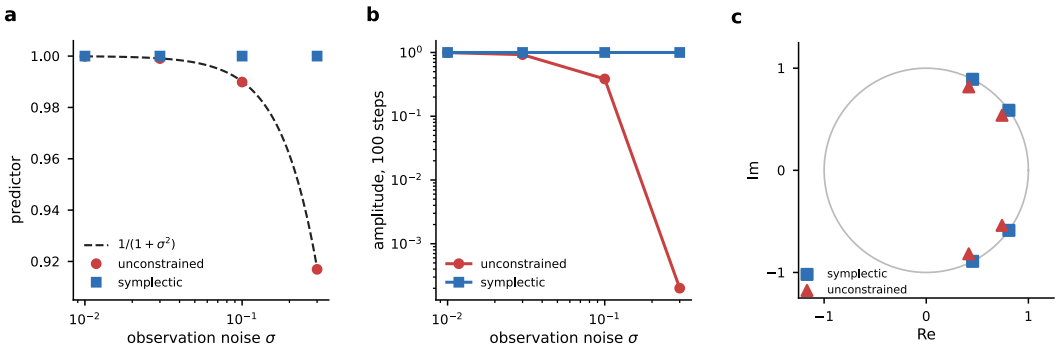


Figure A1: Noise-induced contraction of non-symplectic propagators. **(a)** Predictor $|\mathrm{eig}|$ vs. noise σ follows $1/(1+\sigma^2)$ for the unconstrained fit and remains at 1 for the symplectic fit. **(b)** The 100-step predicted amplitude decays for the unconstrained map. **(c)** At $\sigma = 0.3$, the unconstrained eigenvalues lie inside the unit circle, whereas the symplectic eigenvalues remain on it.

Remark A.23: ellipticity under perturbation. Ellipticity is preserved under symplectic perturbation when the unit eigenvalues have definite Krein signature. This is strong stability in the sense of Gelfand-Lidskii and Krein. We use the definiteness criterion qualitatively and refer to the classical literature for the signature theory.

Coordinate-identifiable dissipation. Let P be a C^1 diffeomorphism, and use the finite-orbit contraction readout $\widehat{\gamma}_n$ defined in (5).

Theorem A.24: gauge invariance under bounded distortion. Let φ be a C^1 diffeomorphism with $\sup_z \left|\log|\det \mathrm{D}\varphi(z)|\right| \le K < \infty$ (bounded distortion), and let $\widetilde{P} = \varphi \circ P \circ \varphi^{-1}$ be the conjugated predictor. Then for every orbit,

$$\left|\widehat{\gamma}_n(z_0; P) - \widehat{\gamma}_n(\varphi(z_0); \widetilde{P})\right| \le \frac{2K}{n\,m} \xrightarrow[n\to\infty]{} 0.$$

If the limit defining γ exists in one coordinate system, it exists and has the same value in every coordinate system satisfying the bounded-distortion condition.

Proof. By the chain rule, $\log\left|\det \mathrm{D}\widetilde{P}(\varphi(z))\right| = \log|\det \mathrm{D}P(z)| + \log|\det \mathrm{D}\varphi(Pz)| - \log|\det \mathrm{D}\varphi(z)|$. Summing along the orbit, the φ-terms telescope to $\log|\det \mathrm{D}\varphi(z_n)| - \log|\det \mathrm{D}\varphi(z_0)|$, bounded by $2K$. □

Lemma A.25: bi-Lipschitz invariance of the Lyapunov spectrum. Let $\widetilde{P} = \varphi \circ P \circ \varphi^{-1}$ with φ a diffeomorphism of uniformly bounded derivative, $\|\mathrm{D}\varphi(z)\| \le B$ and $\left\|\mathrm{D}\varphi^{-1}(z)\right\| \le B$ for all z. Then, whenever the Lyapunov exponents along an orbit of P exist, those along the conjugate orbit of $\widetilde{P}$ exist and coincide.

Proof. Along conjugate orbits the tangent cocycles satisfy $\mathrm{D}\widetilde{P}^{\,n}(\varphi z_0) = \mathrm{D}\varphi(z_n)\,\mathrm{D}P^{\,n}(z_0)\,\mathrm{D}\varphi(z_0)^{-1}$. Taking k-th exterior powers, $\bigwedge^k \mathrm{D}\widetilde{P}^{\,n}(\varphi z_0) = \bigwedge^k \mathrm{D}\varphi(z_n) \bigwedge^k \mathrm{D}P^{\,n}(z_0) \bigwedge^k \mathrm{D}\varphi(z_0)^{-1}$; since $\left\|\bigwedge^k \mathrm{D}\varphi\right\| \le \|\mathrm{D}\varphi\|^k \le B^k$ (and likewise for φ^{-1}),

$$B^{-2k} \left\|\bigwedge^k \mathrm{D}P^{\,n}(z_0)\right\| \le \left\|\bigwedge^k \mathrm{D}\widetilde{P}^{\,n}(\varphi z_0)\right\| \le B^{2k} \left\|\bigwedge^k \mathrm{D}P^{\,n}(z_0)\right\|.$$

The product of the top k singular values is $\left\|\bigwedge^k \mathrm{D}P^{\,n}\right\| = \prod_{i\le k} \sigma_i(\mathrm{D}P^{\,n})$. Applying $\frac{1}{n}\log(\cdot)$ to the bounds above makes the factors $B^{\pm 2k}$ contribute $O(1/n)$, which vanishes as $n \to \infty$. Thus the sums of the top k Lyapunov exponents agree for P and $\widetilde{P}$ whenever they exist. Subtracting consecutive sums gives $\lambda_k(P) = \lambda_k(\widetilde{P})$ for every k. □

Example A.26: unbounded distortion absorbs dissipation. In action-angle coordinates (I, θ) a uniform contraction acts as $I \mapsto e^{-2\gamma} I$, $\theta \mapsto \theta + \mu$. The chart $\rho = \log I$ conjugates it to the translation $\rho \mapsto \rho - 2\gamma$, which is volume-preserving. The chart has $|\det \mathrm{D}\varphi| \to \infty$ toward the fixed point: distortion is unbounded, Theorem A.24 does not apply, and a free encoder can absorb the entire dissipation into its gauge, as observed empirically. Bounded log-volume distortion is therefore sufficient for invariance. Without this condition, or an alternative sublinear-growth condition on the boundary term, the contraction rate may change under reparametrization.

Proposition A.27: the conformal class. Let $M^\top \Omega_Z M = e^{-c}\, \Omega_Z$ with $c \geq 0$ (linear conformal-symplectic). Then: (a) the spectrum is closed under $\lambda \mapsto e^{-c}/\lambda$, so eigenvalues occur in pairs with $|\lambda_i \lambda_i'| = e^{-c}$; (b) $|\det M| = e^{-mc/2}$, so each conjugate pair contracts area by e^{-c} and the per-coordinate contraction rate is $\gamma = c/2$ (equivalently $c = 2\gamma$), monotone and without hyperbolic blow-up when the pairs are balanced on the circle of radius $e^{-c/2}$; (c) the anti-collapse guarantee of Theorem A.10 extends to the conformal class on the encoder side: if the encoder induces ω_Z up to a nonzero scalar factor, nondegeneracy is preserved (a nonzero multiple of a nondegenerate form is nondegenerate) and Theorem A.10 applies verbatim, so full rank of the encoder is still forced. Conformal symplecticity permits a predictor contraction rate of $c/2$, but it does not constrain encoder rank. The encoder-side nondegeneracy condition must therefore be imposed separately.

Proof. (a) As in Lemma A.21, $M^{-1} = e^c\, \Omega_Z^{-1} M^\top \Omega_Z$, so $e^{-c} M^{-1}$ is similar to $M^\top$; if $\lambda \in \mathrm{spec}(M)$ then $e^{-c}/\lambda \in \mathrm{spec}(M)$. (b) Taking determinants in the defining relation, $(\det M)^2 = e^{-mc}$, so $|\det M| = e^{-mc/2}$ and $\gamma = -\frac{1}{m} \log |\det M| = c/2$. (c) is immediate. □

B External-benchmark validation

Figure A2 gives the per-system results summarized in the main text.

C Extended related work

Related studies address representation collapse, long-horizon prediction, structure-preserving dynamics, and system identification.

Collapse and drift in joint-embedding models. Non-contrastive self-supervision uses several mechanisms to prevent collapse. BYOL uses a moving-average target (Grill et al., 2020), and SimSiam isolates the role of stop-gradient (Chen & He, 2021). DINO combines a momentum teacher with centering and sharpening (Caron et al., 2021). Barlow Twins and VICReg regularize cross-correlation or embedding statistics (Zbontar et al., 2021; Bardes et al., 2022). LeJEPA (Balestriero & LeCun, 2025) shapes the full embedding distribution with SIGReg. LeWorldModel (Maes et al., 2026) and Sub-JEPA (Zhao et al., 2026) extend distributional constraints to end-to-end world models. Our analysis combines constraints on the latent distribution and predictor dynamics (§G). LeJEPA-style theory studies identifiability of the latent marginal (Klindt et al., 2026). Here we study coordinate invariants of the latent map.

Long-horizon world modeling. PlaNet introduced multi-step latent overshooting for pixel-based planning (Hafner et al., 2019), Dreamer learns behavior through recurrent latent imagination (Hafner et al., 2020), and MuZero plans with a learned task-relevant model (Schrittwieser et al., 2020). Recent work addresses prediction error by changing either the training horizon or the inference procedure. DPWM optimizes endpoint prediction directly and avoids recursive inference (Li et al., 2026); HorizonDrive trains an autoregressive teacher to recover from prediction-corrupted histories (Zhang et al., 2026); Diffusion Forcing supports variable-horizon video generation beyond the training horizon (Chen et al., 2024); and TD-MPC2 plans over short latent predicted trajectories within model-predictive control (Hansen et al., 2024). DreamerV3 demonstrates recurrent latent imagination across diverse control domains (Hafner et al., 2025), PcLast studies representations designed explicitly for planning (Koul et al., 2024), and autoregressive video world models scale through token prediction or visual-change representations (Wu et al., 2024; Ren et al., 2025). The present analysis addresses the complementary case of recursive prediction by diagnosing predictor geometry and constraining its spectral class.

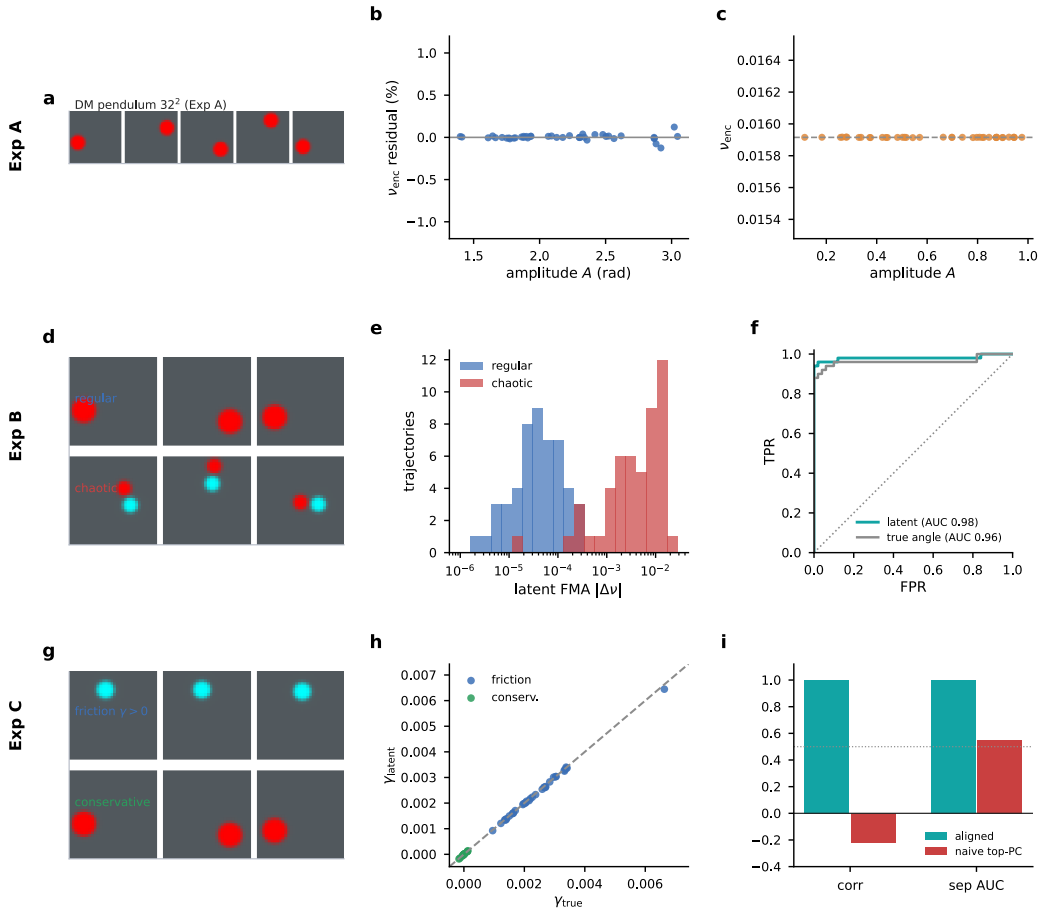


Figure A2: Dynamical diagnostics on the DeepMind Hamiltonian Dynamics Suite (Botev et al., 2021), computed from the frozen V-JEPA 2 latent. **(a to c)** Frequency recovery: residual about the exact pendulum dispersion (median 0.011%) and an amplitude-independent mass-spring control. **(d to f)** Regular versus chaotic motion: frequency-diffusion ratio $112\times$ and AUC 0.98 from latent PCA, compared with 0.96 from a true angle. **(g to i)** Dissipation and observable dependence: damping-rate correlation 0.9995 and AUC 1.00 for the aligned probe, compared with AUC 0.55 for the top principal component, which follows the randomized bob colour.

Structure-preserving predictors. Hamiltonian and symplectic networks (HNN (Greydanus et al., 2019), SympNets (Jin et al., 2020)) supply our predictor classes. Related work imposes explicit mechanical constraints (Finzi et al., 2020) or directly learns exactly symplectic discrete maps with controlled long-time error (Chen & Tao, 2021). Hamiltonian Generative Networks (Toth et al., 2019) and Lagrangian models learned from images (Zhong & Leonard, 2020) jointly learn structured coordinates and dynamics, while Dissipative SymODEN (Zhong et al., 2020) extends the structured class to dissipative, controlled systems. These methods train the representation end to end, often with a decoder, whereas our setting imposes structure in a frozen latent. Symplectic reduced-order models (Chen et al., 2025) likewise learn structure-preserving latents with a decoder. Contractive world models (Lu et al., 2026) stabilize predicted trajectories by pulling the spectrum inward; our predictor attenuation analysis characterizes the potential tradeoff when contraction exceeds the physical dissipation.

Operator-theoretic learning and system identification. Equation (1) has the form of a learned Koopman operator fitted by extended dynamic mode decomposition (Takeishi et al., 2017; Lusch et al., 2018; Otto & Rowley, 2019; Azencot et al., 2020). Additive sensor noise gives the population limit $A\Sigma(\Sigma+\sigma^2 I)^{-1}$ and hence the shrinkage described by our predictor attenuation analysis (Dawson et al., 2016). Noise-covariance subtraction, forward-backward DMD (Dawson et al., 2016), and

TLS-DMD (Hemati et al., 2017) remove this classical attenuation. In a recursive JEPA prediction, the factor compounds as $(1+\sigma^2)^{-n}$. Statistical correction removes the bias, whereas a symplectic parameterization additionally imposes a reciprocal spectrum. Lemma A.17 gives the corresponding log-determinant correction for the contraction rate.

Benchmarks and encoder probing. The Hamiltonian Dynamics Suite (Botev et al., 2021) and SyMetric (Higgins et al., 2021) evaluate models trained on their videos, whereas DINO-WM predicts pretrained visual features for decoder-free planning (Zhou et al., 2025). Frozen visual encoders can support physical judgments (Garrido et al., 2025); here (ν, λ, γ) are read from a frozen natural-video encoder and tested under iteration of its predictor. The analysis combines a gauge-invariant contraction measurement, component-level rank and spectrum criteria, and a dynamical dissection of a frozen large-scale model. Topological obstructions to flat latent charts connect this setting to homeomorphic representation learning (Falorsi et al., 2018).

D Experimental details

The following sections describe the systems, training procedures, and model evaluation setups.

Systems and experiment labels. Table A1 defines the experiment numbers (Exp. 1, 1b, 3, 4, 4b, 7, 13, 14) and the T labels used throughout. The systems progress from controlled analytic dynamics to external videos and a frozen predictor.

Table A1: Validation systems and experiment index.

Level	System	Source	Experiments
T_0	linear symplectic rotation on $\mathbb{R}^4$	authors (analytic)	1, 1b, 1c, 3, 4, 4b, 9
T_1	pendulum and standard map	authors (analytic)	5, 6, 7, 7b to 7d
T_2'	rendered pendulum video with V-JEPA 2	authors (rendered)	13, 14
T_2''	Hamiltonian Dynamics Suite videos	public benchmark	Figure A2
T_2'''	laboratory pendulum video	public recording	§3
T_3	frozen V-JEPA 2 predictor under iteration	public checkpoint	12, 14

The solvable T_0 system. The controlled system is a linear symplectic rotation on $\mathbb{R}^4$ (2 degrees of freedom). Its dynamics are $z_{t+1} = Rz_t$, with R block-diagonal in two canonical planes at tunes $\mu_1 = 0.2\pi$ and $\mu_2 = 0.35\pi$. The tunes are distinct, so Theorem A.8 applies. Analytic values for rank, spectrum, tune, and invariants provide the calibration targets. Observations are the state, optionally rendered to pixels, with i.i.d. isotropic Gaussian noise of standard deviation σ added per frame.

Training. Unless noted, training uses the unregularized one-step latent risk (2). The target is online (non-EMA), with no stop-gradient or anti-collapse term, so protection against collapse comes from the architecture. Configurations combine encoder $\in$ {free MLP, symplectic}, predictor $\in$ {MLP, symplectic}, and anti-collapse $\in$ {none, SIGReg, VICReg}. Each is trained for 3,000 to 5,000 Adam steps over 3 random seeds, in clean and noisy regimes ($\sigma = 0.1$). Reported values are medians across seeds.

Nonlinear probe systems. Two systems with analytic answers extend the linear theory. The pendulum is integrated in closed form via Jacobi elliptic functions, giving the exact amplitude-dependent tune $\nu(\theta_0) = \nu_0\,(\pi/2)/K(\sin^2(\theta_0/2))$ and its small-amplitude $-1/16$ shift. The Chirikov standard map $(\theta, p) \mapsto (\theta + p', p')$ with $p' = p + K\sin\theta$ is a controlled route into chaos: its rotation number is analytic at $K = 0$ and the chaotic fraction grows with K (we use $K = 0.5$ and $K = 1.0$).

V-JEPA 2 evaluation. The public frozen V-JEPA 2-AC checkpoint contains a ViT-g encoder with approximately 1.0B parameters and an action-conditioned predictor with approximately 0.3B parameters. The iterated state is the full one-frame token field, $z_t \in \mathbb{R}^{256\times1408}$, with flat dimension 360,448. Fixed action and proprioceptive conditioning define the autonomous map $z_{t+1} =$

$P(z_t; a^*, s^*)$. A rolling context of k frames defines a closed map on $x_t = (z_{t-k+1}, \ldots, z_t)$. The experiments in §3 and Appendix I fit low-dimensional observables of this state. The single-frame pendulum projection is hyperbolic, while delay embedding and the native clip input admit elliptic fits.

Encoding uses bf16, and the Benettin QR and finite-difference calculations use fp32. All 192 dropout modules are inactive. Two predictions from the same initial latent are bitwise equal over 512 steps. A 10^{-4} perturbation grows by $83\times$ before saturating on a bounded attractor. The leading exponent is stable over two decades of finite-difference scale, with $\lambda_1 = 0.146, 0.145, 0.151$. Repeating the 150-step calculation with bf16 predictor matrix products gives $\lambda_1 = 2.09$ instead of 0.145. The relative perturbation $\epsilon_{\text{rel}} = 10^{-3}$ is below the approximately 4×10^{-3} bf16 relative resolution. The fp32 calculation removes this $14\times$ roundoff inflation.

All estimated exponents remain positive with rolling contexts of $k = 2$ and $k = 4$ frames. Changing the initial condition from the robot clip to the real pendulum footage gives $\lambda_1 = 0.043$ to 0.120. Across the tested initial conditions and context lengths, λ_1 spans 0.04 to 0.16.

SIGReg port. SIGReg (Balestriero & LeCun, 2025) was designed for image representations without a temporal predictor. Its penalty is applied to the encoder output in the temporal prediction objective, using the authors' official single-GPU implementation. The resulting latent statistic is used as an isotropy score.

Prediction fidelity and the horizon τ. At step k, prediction fidelity is the cosine similarity $\cos(P^{\circ k}(z_0), E(x_k))$ between the flattened predicted latent and encoded future. The floor, defined as the median over the last 16 steps, is 0.48 to 0.51 across amplitudes. The horizon τ is the first step at which the similarity falls halfway from its one-step value to this floor. Figure 1d gives $\tau = 21, 15, 15$ for $\theta_0 = 0.4, 1.0, 1.6$. This latent-space metric accompanies the physical observation that the injected frequency enters the model's intrinsic band.

The ν-probe. A ridge-regression probe maps the frozen latent to signed oscillator angle with held-out $R^2 = 0.995$ on V-JEPA 2. Frequency-map analysis is then applied to the decoded angle. A pooled component instead selects the second harmonic or rendered colour (§2).

E Diagnostics toolkit and reproduction

This section defines the three dynamical instruments and their calibration on a learned model.

Frequency-map analysis (FMA). For a complex observable $f(z_t) = u(z_t) + i\, v(z_t)$, a Hann-windowed discrete Fourier transform locates the dominant-frequency bin (Laskar, 1993). Maximizing $\left|\sum_t f(z_t) w(t) e^{-i2\pi\nu t}\right|$ near that bin gives sub-bin precision. The difference $|\Delta\nu|$ between the two trajectory halves measures frequency diffusion.

Lyapunov spectrum (Benettin QR). The tangent vectors evolve as $V_{t+1} = \mathrm{D}P(z_t)V_t$ and are periodically re-orthonormalized by $V_t = Q_t R_t$. The exponents are $\lambda_i = \lim_{T\to\infty} T^{-1} \sum_{t<T} \log[R_t]_{ii}$. Jacobian products use finite differences for V-JEPA 2 and automatic differentiation for the toy predictors. Appendix D gives the numerical precision and finite-difference settings.

Local stretching and contraction rate. The singular values of $\mathrm{D}P$ give the per-direction local stretching and $\log|\det \mathrm{D}P|$ the local log-volume change. These local quantities are not coordinate-invariant. Their asymptotic orbit average gives the coordinate-identifiable contraction rate $\gamma = -\frac{1}{m}\, \mathbb{E}_t \log|\det \mathrm{D}P(z_t)|$ (cf. (5)), which is invariant under bounded-distortion changes of latent gauge (Theorem A.24). It is estimated from the Jacobian of a free predictor because the latent marginal does not identify temporal contraction.

Operating-range calibration. Each measurement is calibrated by varying the one-step latent fit error and recording the threshold at which the target result is retained. For example, the pendulum tune-shift ordering is recovered at Spearman ≥ 0.9 down to a one-step latent MSE of $\sim 10^{-3}$. This

threshold depends on the system and is stricter for resonant systems such as the standard map than for smooth systems such as the pendulum. The calibration is repeated for each target model.

Latent topology. The standard map lives on a cylinder, whereas a flat $\mathbb{R}^2$ latent represents its winding angle by an unbounded coordinate. In the small-capacity model, the nonlinear-symplectic layers then saturate to nearly constant kicks and the prediction drifts. Applying FMA to the true orbit after encoding localizes the error to the representation, with a folding-signature rank correlation of approximately -0.3. A polar chart embeds the cylinder as a bounded annulus and reduces the divergence by approximately $500\times$. Larger models are required to determine whether this topology-structure tradeoff persists with increased capacity.

Reproduction. Every experiment is deterministic given its seed. The T_0 studies use 3 seeds and report medians. The V-JEPA 2 analysis uses the public frozen checkpoint with fixed action and proprioceptive conditioning. Figure 3 reports the Lyapunov spectrum, its sensitivity to the finite-difference scale, frequency diffusion, and harmonic selection. Table A4 lists the measurements for each system, using the definitions above.

F Empirical validation of the structural implications

Table A2: The four structural implications and their empirical validation at a glance.

Topic	Content	Validation	Key number
reduction	min. dim. $= 2\times$DOF; pairs inseparable	Exp. 1b	symplecticity 0.97 vs 0.43
encoder	symplectic encoder prevents collapse	Exp. 1	rank $4/4$, cond 1.08
predictor	noise $\Rightarrow$ artificial damping $1/(1+\sigma^2)$	Exp. 3,7	$\lvert\text{eig}\rvert \to 0.990$; vol $\to 0.018$
dissipation	log-det gives a coordinate invariant	Exp. 4,4b	whitening 191% / log-det 0%

Each statement of §2 is tested on the solvable T_0 system, a block-diagonal symplectic rotation with analytic ground truth, and on its nonlinear extensions. Table A4 collects the instrument measurements.

Table A3: Structure diagnostics on the unregularized solvable system. Rank: numerical rank of the 4-dimensional latent covariance. Symplecticity: $\min \text{sv}(V^\top \Omega V)$ (1 = canonical pair preserved). Artificial damping: $1 - |\text{eig}|$ at $\sigma \approx 0.1$ (0 = on the unit circle). N/A marks quantities not measured.

Configuration	rank $(/4)$	symplecticity	artificial damping $1 - \lvert\text{eig}\rvert$ (%)
free enc. / MLP (least squares)	2	0.43	0.99
free enc. / symplectic P	2	0.97	0.00
+ SIGReg	4	N/A	1.03
+ VICReg	4	N/A	1.03
symplectic enc.	4	1.00	0.00
Coordinate-identifiable dissipation, true $\gamma = 0.030$:			
whitening gauge	$\hat{\gamma} = 0.087$	$(+191\%$ error)	
log-det invariant	$\hat{\gamma} = 0.030$	(exact, 0%)	
symplectic-P control	$\hat{\gamma} = 0.000$	(conservative)	

Lossless latent reduction: canonical-pair inseparability (Exp. 1b). When a symplectic predictor's latent partially collapses (rank $4 \to 2$), the retained 2D subspace is a canonical symplectic plane. Its restricted symplectic form has minimum singular value 0.973 ± 0.012, and the learned propagator matches the induced $M = CRC^+$ to relative error below 10^{-2}. An unconstrained predictor instead collapses onto a Lagrangian or mixed subspace with symplecticity 0.43 ± 0.31, whereas the symplectic encoder remains full rank with symplecticity 1.00 (§G). In the frozen video latent, the single-frame projection gives a hyperbolic fit with $|\lambda| \approx 1.2$, while the native multi-frame clip gives an elliptic fit with $|\lambda| = 1.00$ (Figure A5). This spectral difference is compatible with temporal context supplying a phase-space observable.

Encoder-side nondegeneracy (Exp. 1). Under the unregularized objective, the symplectic encoder retains full numerical rank $(4/4)$ with condition number 1.08. The free encoder collapses to rank 2 with either predictor, with a condition number of approximately 10^{29}. The symplectic encoder's latent covariance has spectrum $[1.05, 1.01, 0.99, 0.97]$ (§G). Across seeds, these results agree with the implication from a nondegenerate symplectic form to $\mathrm{rank}C = m$.

Predictor attenuation and stability (Exp. 3, with two nonlinear confirmations). At $\sigma = 0.1$, the unconstrained eigenvalue magnitude is 0.98989, compared with the predicted 0.99010, and the 100-step predicted volume falls from 1 to 0.018. At $\sigma = 0.3$, the volume reaches 10^{-15}. The symplectic fit remains on the unit circle to 2×10^{-8} and preserves unit volume. Volume preservation alone does not ensure stability: a nonlinear symplectic network grows in norm by $344\times$, and a hyperbolic linear-symplectic layer grows by $10^6\times$ (Figure A1). With $N = 2 \times 10^4$ pairs and three seeds, TLS-DMD, forward-backward DMD, and noise-corrected DMD restore the mean eigenvalue magnitude to within 6×10^{-4} of unity over $\sigma = 0.03$ to 0.3. Statistical de-biasing therefore removes the shrinkage, whereas the structural constraint additionally guarantees a reciprocal spectrum. In dissipative systems, MEND obtains the contraction rate from the independent coordinate-identifiable measurement (Table 1).

Coordinate-identifiable dissipation (Exp. 4, 4b). On dissipative data with true $\gamma = 0.03$, encoder gauge-fixing by whitening gives 191% relative error, whereas the finite-time orbit-averaged log-determinant of a free predictor recovers γ to 0.0%. A strict symplectic predictor attains loss 3×10^{-7} by absorbing the contraction into encoder distortion, with latent-covariance condition number 55, and therefore does not represent γ. The same effect appears with a full distributional regularizer (§G): forcing isotropy gives 155% error when γ is read through a conformal predictor with a learned scale factor, compared with 2.4% for the free-predictor log-determinant. On the external benchmark, the latter readout recovers the damping rate to 1.3% (Figure A2).

G GEOMETRY VS. DISTRIBUTION: FULL COMPARISON

Distributional and geometric constraints act on different quantities and can be combined. SIGReg (Balestriero & LeCun, 2025) shapes the embedding distribution toward an isotropic Gaussian, and VICReg (Bardes et al., 2022) constrains its first two moments. The geometric conditions constrain the map (E, P). Experiments on the solvable T_0 system combine encoder $\in$ {free, symplectic}, predictor $\in$ {MLP, symplectic}, and anti-collapse $\in$ {none, SIGReg, VICReg}. The tests use the unregularized objective, clean data and $\sigma = 0.1$, and three seeds. Both types of constraint prevent collapse, while predictor geometry also controls the spectrum. Symplectic-predictor configurations keep $|\mathrm{eig}(P)| = 1.0000$. The tested SIGReg- and VICReg-only configurations contract to $1/(1 + \sigma^2)$ under noise (Figure 2a). Combining a free encoder, a symplectic predictor, and SIGReg preserves full rank and $|\mathrm{eig}| = 1$.

The comparison uses two nondegeneracy theorems and a dissipative example. Table A4 collects the measurements, and Figure A3 compares rank and isotropy.

Nondegeneracy conditions for the distribution and encoder. The distributional and geometric routes use nondegeneracy conditions on different objects. Matching finitely many moments does not determine a distribution (LeJEPA, Thm. 3), so VICReg can whiten the covariance while leaving the latent non-Gaussian. Configuration F has a SIGReg-of-latent score of 25, an order of magnitude above the ≈ 1 values of D and E, despite its isotropic second moment. SIGReg instead matches the full characteristic function. For the geometric route, nondegeneracy of the symplectic form makes the encoder a full-row-rank submersion: its restriction to $(\ker C)^\Omega$ is an isomorphism onto $\mathbb{R}^m$, so C cannot be rank-deficient by the encoder-side nondegeneracy condition. The measured encoder remains rank 4 with condition number 1.1.

Dissipative dynamics (Exp. 4b). When the true dynamics contract, isotropy of the latent marginal does not determine the contraction rate. On dissipative T_0 data with true $\gamma = 0.03$, a conformal predictor with a learned scale factor gives 155% error under SIGReg. The whitening gauge discussed under coordinate-identifiable dissipation gives 191% error. With SIGReg, the latent isotropy score is 1.7; without it, the score ranges from 88 to 1600. The finite-time orbit-averaged log-determinant

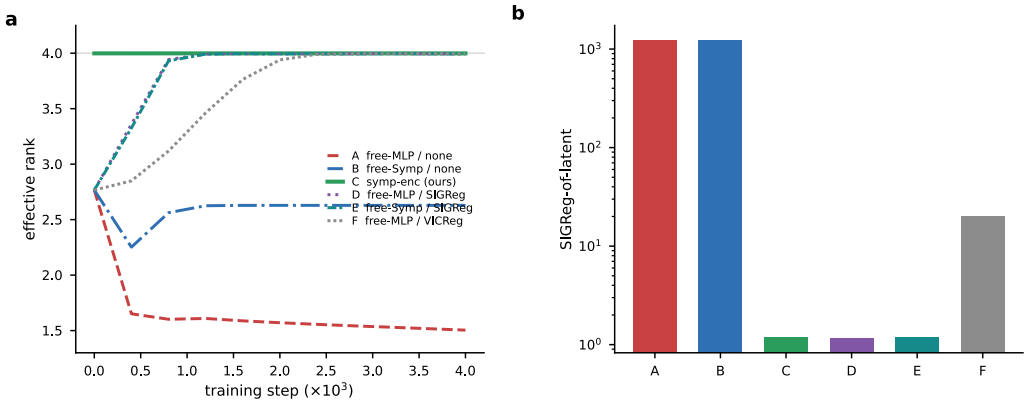


Figure A3: Geometry vs. distribution on the unregularized objective. **(a)** Effective-rank curves for the six configurations: A/B collapse, whereas C/D/E/F remain at rank 4. **(b)** SIGReg-of-latent (isotropy score, lower is better): the SIGReg configurations D/E and the symplectic encoder C are isotropic, with covariance spectrum $\approx [1.05, 1.01, 0.99, 0.97]$ for C; VICReg configuration F is not.

of a free predictor gives 2.4% error on the same data. The predictor Jacobian measures temporal contraction, which the latent marginal does not identify.

Table A4: Diagnostics across systems. The top block compares geometry and distribution on the two-degree-of-freedom T_0 system with an unregularized objective and three seeds; the bottom block applies the same instruments to nonlinear systems, benchmarks, and V-JEPA 2. Anti-collapse is measured by numerical rank (↑; full rank $= 2\times\text{DOF} = 4$) and SIGReg-of-latent (↓; ≈ 1 denotes an isotropic Gaussian). Dynamics are measured by the predictor spectrum at $\sigma = 0.1$, frequency recovery, and a chaos or dissipation readout. The best T_0 configuration is shown in **bold**; N/A denotes an inapplicable measurement.

		anti-collapse		spectrum	frequency	
System / config	probe	rank	SIGReg-lat	|eig|	recovery	chaos / dissipation
Solvable T_0 (2 DOF): encoder×predictor×anti-collapse, six configs A to F						
A free MLP / none	N/A	2.3	1225	0.689	err 0.247	collapses (rank 2)
B free Symp / none	N/A	2.0	1227	1.000	err 0.013	collapses (Lagrangian)
C **symp. enc. / intrinsic**	N/A	**4**	**1.2**	**1.000**	**err** 0.000	survives, no damping
D free MLP / SIGReg	N/A	4	1.2	0.990	err 0.000	survives, still damps
E free Symp / SIGReg	N/A	4	1.2	1.000	err 0.005	survives
F free MLP / VICReg	N/A	4	20.4	0.990	err 0.000	survives, still damps
Nonlinear systems, benchmarks, and V-JEPA 2 (analytic or frozen-encoder probe)						
pendulum (self-rendered)	ν-probe	N/A	N/A	N/A	coeff $-\mathbf{0.062}$ (vs $-1/16$)	FMA 7×10^{-11}
standard map K=0.5	FMA	N/A	N/A	N/A	rot. # 3.4×10^{-9}	chaotic frac. 0.01
standard map K=1.0	FMA	N/A	N/A	N/A	N/A	chaotic 0.41; $\lvert\Delta\nu\rvert$ 1.7×10^{-6}
DM pendulum	ν-probe	N/A	N/A	N/A	R^2>0.9999; err 7×10^{-5}	N/A
DM mass-spring	ν-probe	N/A	N/A	N/A	err 3.4×10^{-5}	ν amplitude-indep.
DM double pendulum	FMA	N/A	N/A	N/A	N/A	chaos AUC **0.98** (vs 0.96 true)
DM friction pendulum	γ-probe	N/A	N/A	N/A	N/A	γ corr **0.9995**; err 1.3%
V-JEPA 2 probe	ν, λ	N/A	N/A	N/A	R^2=0.995; ∼1%	$\lambda_1\approx0.15$; $\geq$ 12 pos.; $\tau\approx15$ to 21

H REAL-PENDULUM EXPERIMENT: METHODS AND UNCERTAINTY

Footage and units. The experiment uses a public recording of a laboratory pendulum (661 frames, 18.7 periods) from `https://www.teachphysics.org/`. The recording is licensed under CC BY-NC-SA 4.0 and was cropped with the creator's permission. The footage has been re-encoded, and the original capture clock is unavailable, so all quantities are reported per frame or in dimensionless form. A uniform speed error rescales the true and latent frequencies identically and cancels from every comparison.

Pixel ground truth. The bob is tracked by a matched disc filter with a radius measured from the data and sub-pixel centroid refinement. Hough accumulation over the string locates the pivot, with a score contrast of $11\times$ against a displaced control. Two independent trackers, a disc filter and an HSV colour blob, agree to $0.08°$. The bob's pivot distance varies by 0.72% and is independent of the fit. This confirms the planar swing geometry used to measure the angle directly. The timebase check finds zero duplicated frames among fast-moving frames and zero phase-advance flat spots. Phase jitter is $0.88\times$ the level produced by the measured tracking noise alone, indicating no re-timing from transcoding.

Estimator calibration and uncertainty. A noiseless simulation recovers γ_{true} to eleven digits for every lag and averaging length. The one-frame delay embedding has condition number 11, compared with 1.03 at a quarter-period delay. The one-step determinant overestimates the physical contraction rate by a factor of 12. Direct regressions give $\widehat{\gamma}_n = 4.79, 2.83, 1.75, 0.90, 0.41, 0.37 \times 10^{-3}$ at $n = 1, 2, 4, 9, 70, 140$ and 3.49×10^{-4} at $n = 420$. The pixel-tracking reference is 3.84×10^{-4}. The decrease with lag follows the $1/n$ bias of Lemma A.17. On the pixel-tracked angle, $\widehat{\gamma}_n$ remains within 0.9% of the reference for lags of 18 to 210 frames. This bounds the effect of amplitude-dependent frequency drift over those lags.

A probe fitted on the first 60% of the clip gives an envelope rate 43% below the pixel value. The full-clip probe and model-free envelope fit agree at a 9% offset. Confidence intervals use blocks of one period. Figure A4 compares six predicted trajectories.

De-biasing baselines for predictor attenuation. The fitted linear predictors in Table 1 are 2×2 one-step maps in the delay embedding of the probe observable. Each map is iterated autonomously for 651 steps, and the observed probe trajectory provides the reference. The de-biased and structured predictors remain on the observed manifold. The de-biased contraction rates are 25% below the pixel value. The independently measured rate supplied to the conformal predictor is 9% below it. Noise-corrected DMD depends on the chosen noise variance σ^2. The n-step log-determinant excess gives $\sigma^2 = 0.44\%$ ($\sigma = 6.6\%$) through Lemma A.17. The probe's held-out residual $1 - R^2$ gives $\sigma^2 = 2.4\%$ ($\sigma = 15.5\%$). The two corrections give either a 62% overestimate or divergence. The residual estimate is larger because it includes model and tracking error alongside the i.i.d. observation noise assumed by the estimator.

Terminal-amplitude estimator. We compare three estimates of terminal amplitude. The first is the ratio of a fitted exponential envelope over the peak span, used throughout for predictions and the observed reference. The second uses the Hilbert envelope averaged over the first and last 40 frames. The third divides the last peak by the fitted envelope. They give 0.82, 0.79, and 0.89 for the observed trajectory, and 0.82, 0.78, and 0.82 for the conformal prediction. The terminal amplitude varies by approximately $\pm 8\%$ across these estimators.

Interpreting 0.82 as exponential decay over all 661 frames gives $Q \approx 290$, compared with $Q \approx 231$ from pixel tracking. Three measured effects account for the difference. First, the probe observable decays about 10% more slowly than the pixel angle. The same envelope fit gives rates of 3.73×10^{-4} and 4.13×10^{-4} per frame, consistent with the probe error reported for γ. Second, the peak span is shorter than the full clip. Third, the envelope-fitted rate for the pixel angle is 7% above the tracker's γ_{true}.

The exponential fit has $R^2 = 0.68$ for the pixel log-envelope and 0.50 for the probe. The $1/t$ law for quadratic drag gives $R^2 = 0.58$ and 0.43, respectively. At $Q \approx 231$, the decay over 18 periods is too small to distinguish these envelope shapes. The constant-γ conformal model therefore gives a first-order description over the observed interval.

Anisotropy of the errors-in-variables attenuation. The estimate $\sigma = 6.0\%$ from $|\mathrm{eig}| = 0.996072$ uses the scalar isotropic law in Theorem A.15. We also evaluate the anisotropic expression $R\Sigma(\Sigma + \sigma^2 I)^{-1}$ from the following remark. The measured 2×2 covariance has eigenvalues 0.969 and 1.029, with an anisotropy ratio of 1.06. The two attenuation eigenvalues are 0.99545 and 0.99571, compared with the isotropic value 0.99561. The maximum deviation is 0.02%, so the scalar law is accurate to two parts in 10^4 for this embedding.

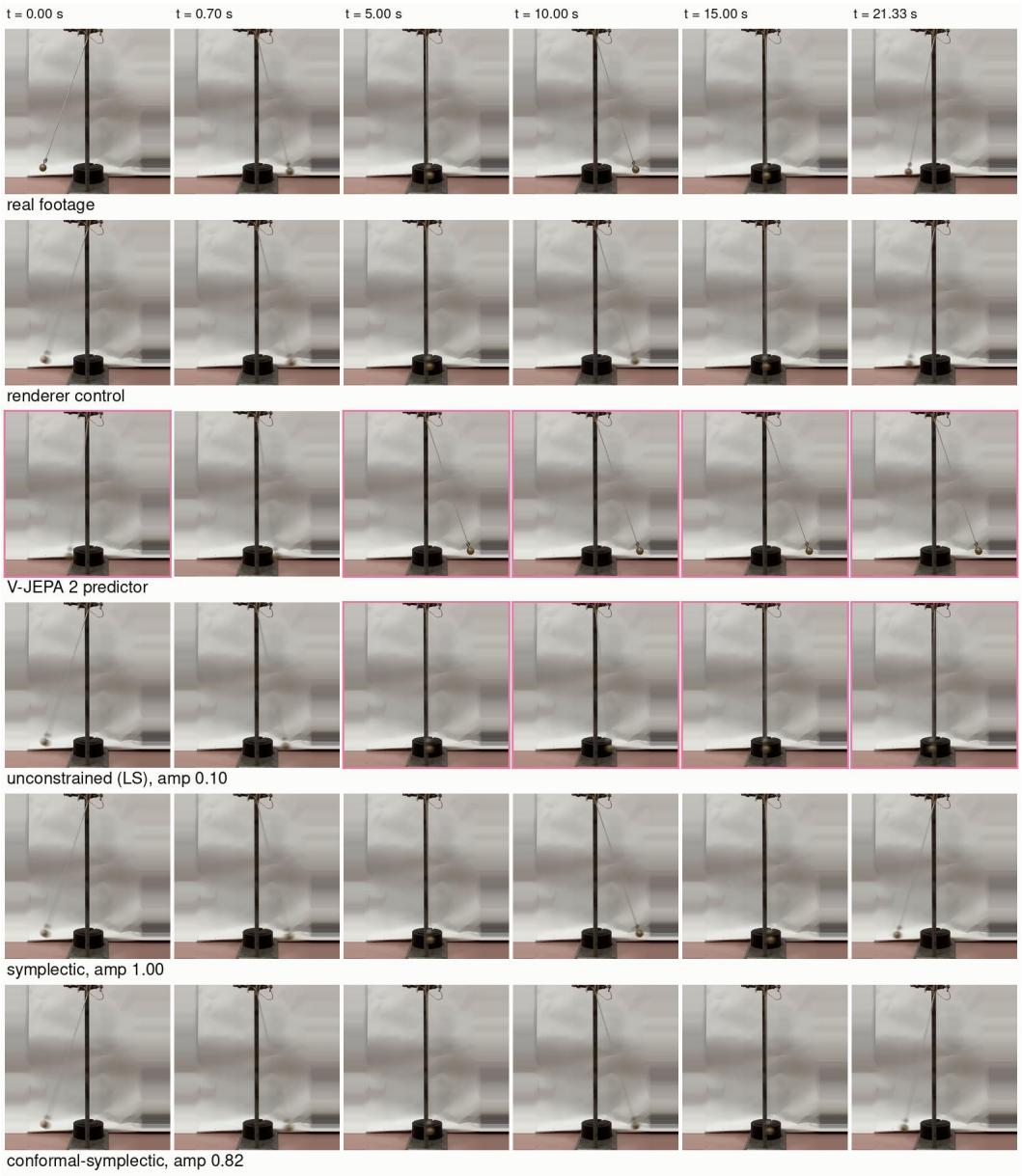


Figure A4: Six predicted trajectories in the external experiment, shown at six times. From top to bottom: real footage; a renderer control reconstructing the true trajectory; V-JEPA 2's predictor; unconstrained least squares; symplectic prediction; conformal-symplectic prediction. Every row uses the same nearest-frame renderer, which retrieves appearance from the clip. The renderer control bounds the rendering error. Red borders mark phase states absent from the observed trajectory. Retrieval clamps to observed frames, so the unconstrained prediction can appear plausible even when its phase state is out-of-distribution. The borders identify this departure. For V-JEPA 2, the decoded angle is a linear functional that fails when evaluated outside its fitted domain. Footage: `teachphysics.org` (CC BY-NC-SA 4.0), cropped, used with permission.

I Additional figures

Figure A5 compares single-frame and multi-frame phase portraits to test the phase-space observable requirement.

Requirements for structure-preserving prediction. A structure-preserving predictor requires compatible dynamics and observables. The elliptic reciprocal prior applies to conservative dynamics. Anisotropic dissipation requires per-mode spectral structure beyond the uniform conformal class in coordinate-identifiable dissipation. Lossless latent reduction also requires a phase-space observable containing each canonical pair (q, p) (lossless latent reduction). In Figure A5, the single-frame projection forms a folded arc and gives a hyperbolic fit with $|\mathrm{eig}| > 1$. The native multi-frame projection forms a closed orbit and gives $|\mathrm{eig}| = 1$. A learned canonical map from frozen video latents to (q, p) remains to be developed.

Effect of encoder pretraining. We test whether the elliptic fit depends on encoder pretraining by repeating the Figure A5 experiment with randomly initialized weights and three seeds. The four libration trajectories and fitting procedure are unchanged. The inputs use $K = 8$ clips at stride 2, followed by mean-pooling, projection to four dimensions by SVD, and a symplectic fit. The multi-frame input gives $|\lambda| = 1.000$ for every tested trajectory and seed, as it does for V-JEPA 2.

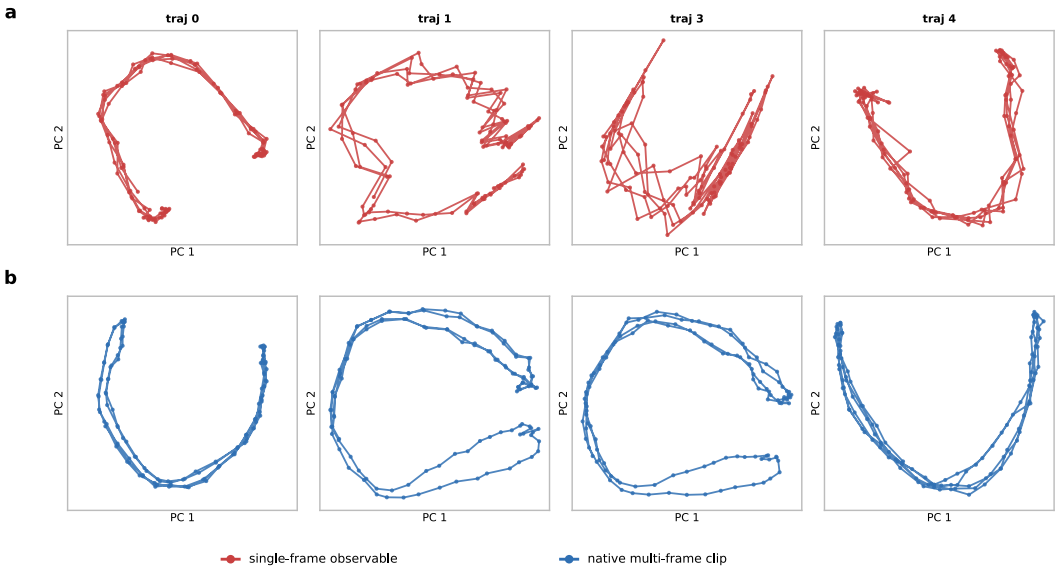


Figure A5: **Observable requirement for lossless latent reduction.** Low-dimensional PCA phase portraits of a frozen V-JEPA 2 latent along four conservative pendulum trajectories. **(a)** The single-frame projection forms an open arc and gives a hyperbolic symplectic fit, $|\lambda| \approx 1.2$ and at most 1.25. **(b)** The native multi-frame projection forms a closed orbit and gives an elliptic fit, $|\lambda| = 1.00$.

We compare prediction error against persistence using the skill $1 - \mathrm{MSE}(\mathrm{fit})/\mathrm{MSE}(\mathrm{identity})$. An identity prediction has zero skill. For the random encoder, the clip fits have median skill 0.39 to 0.76 across three seeds, compared with 0.217 for V-JEPA 2. The elliptic clip fit therefore arises without pretraining. Its variation across trajectories is larger, spanning 0.04 to 0.79, compared with 0.150 to 0.331 for V-JEPA 2.

For single-frame inputs, the random encoder gives $|\lambda| = 1.01$ to 1.04 and skill values of 0.007 to 0.019, compared with 0.071 for V-JEPA 2. These fits perform close to persistence, so their near-unit eigenvalues do not explain the hyperbolic fit for V-JEPA 2. The V-JEPA 2 clip fit in Figure A5b is elliptic, with a prediction skill of 0.217 over persistence.