

Skeptical review: Symplectic Emulation of N-body Dynamics with Hamiltonian Graph Neural Networks

Summary

This paper proposes a Symplectic Hamiltonian Graph Neural Network emulator for softened gravitational N-body dynamics. Instead of learning a direct state-to-state map, the method learns a separable Hamiltonian $H(\mathbf{q}, \mathbf{p}) = T(\mathbf{p}) + U(\mathbf{q})$, where T is analytic and $U(\mathbf{q})$ is parameterized as a permutation-invariant sum over pairwise interactions computed by an MLP on softened distances (Sec. 2.2, Eq. (1)–(3)). Forces are obtained by automatic differentiation, ensuring a conservative force field. A differentiable leapfrog (kick–drift–kick) integrator is unrolled in training so the learned discrete-time flow is symplectic by construction for the learned separable Hamiltonian (Sec. 2.3). Experiments train on $N = 50$ virialized Plummer spheres generated with softened gravity and leapfrog integration, using rotation/translation augmentation and a two-stage radial curriculum that initially masks the dense core (Sec. 2.1, Sec. 2.3). Evaluation reports trajectory reconstruction error and tests for physically meaningful behavior (bounded energy error, time reversibility, and a numerical Jacobian-determinant-based volume check) and includes zero-shot transfer to $N = 25$ and $N = 100$ (Sec. 3.1–3.2). The overall direction is well-motivated, but the manuscript currently underspecifies the architecture/training and relies on weak baselines and largely qualitative reporting; several claims (novelty/positioning, symplecticity verification, and breadth of “generalization”) need tightening and/or stronger experiments to be fully convincing.

Strengths

- Well-motivated structural choice: learning a Hamiltonian (rather than a black-box next-step predictor) targets long-horizon stability and physical plausibility for chaotic N-body dynamics (Sec. 1, Sec. 2.2, Sec. 4).
- Conservative dynamics by construction: parameterizing $U(\mathbf{q})$ as a scalar potential and differentiating it to obtain forces guarantees a curl-free force field in \mathbf{q} -space (Sec. 2.2, Eq. (3)).
- Integrator-aware design: embedding a differentiable leapfrog (kick–drift–kick) integrator aligns learning with a standard symplectic, time-reversible method for separable Hamiltonians (Sec. 2.3).
- Permutation-invariant interaction form: the pairwise aggregation in $U(\mathbf{q})$ is naturally invariant to particle reindexing and is a sensible inductive bias for N-body systems (Sec. 2.2, Eq. (1)).
- Evaluation attempts to go beyond pointwise rollout error by checking energy behavior, reversibility, and phase-space volume-related diagnostics (Sec. 2.4, Sec. 3.2).

- Practical training choices are physically informed: softened distances avoid singular gradients (Eq. (2)), and the radial curriculum is tailored to the core-halo structure of Plummer spheres (Sec. 2.1, Sec. 2.3).

Major issues

1. **Baseline comparisons are far too weak to support the paper’s central claims. Sec. 3.1 / Fig. 1 compare mainly against a trivial “point mass at origin” baseline, which is not representative of modern neural emulators or prior structure-preserving approaches. As a result, it is impossible to attribute gains specifically to (i) Hamiltonian parameterization, (ii) conservative forces via autograd, or (iii) symplectic unrolling (Secs. 1, 3.1, 4).**

Recommendation: Expand Sec. 2.4 (metrics/protocol) and Sec. 3.1–3.2 (results) to include a meaningful baseline suite and targeted ablations, run on identical train/test splits and horizons. At minimum include: (a) a non-symplectic state-to-state baseline (MLP/RNN/GNN next-step predictor) integrated with Euler/RK methods; (b) a force-regressing GNN baseline (learn accelerations/forces directly) with the same integrator; (c) an HNN with the same $U(\mathbf{q})$ but trained without symplectic unrolling (e.g., one-step loss or vector-field loss) and/or evaluated with a non-symplectic integrator; and (d) an “integrator ablation” where the learned Hamiltonian is integrated with RK4 vs leapfrog to isolate the role of symplectic integration. Report trajectory error and invariant metrics (Sec. 3.2) for $N = 25/50/100$ across these baselines.

2. **Quantitative reporting is insufficient and often qualitative (“significant,” “exceptionally low,” “ ≈ 1 ”). Figures and text do not provide enough explicit numerical values, uncertainty, or precise metric definitions (Sec. 2.4, Sec. 3.1–3.2; Fig. 1–2). This weakens evidence and prevents reproducible comparison.**

Recommendation: In Sec. 2.4, give explicit mathematical definitions for each metric, including how quantities are aggregated across particles, coordinates, time, and test simulations (e.g., per-particle vs global; mean vs median; whether velocities/momenta are included). In Sec. 3.1–3.2, report summary statistics (mean \pm std or median/IQR) over multiple test realizations and random seeds: trajectory MSE at several times (e.g., $t = 1, 2.5, 5.0$), maximum/RMS $|\Delta H|$ (clearly defined), reversibility error distribution, and the volume/symplectic diagnostic distribution. Add error bars to Fig. 1 and summary bands/quantiles to Fig. 2, and state the number of runs.

3. **Novelty/positioning relative to prior work is not clearly established. The method combines elements common in Hamiltonian Neural Networks, symplectic neural ODE/Symplectic ODE-Nets, and interaction networks/GNN N-body emulators, but the manuscript lacks a clear statement of what is new beyond that combination (Sec. 1–2).**

Recommendation: Add a dedicated Related Work section (Sec. 1 or Sec. 2.x) covering HNNs, symplectic neural integrators / Symplectic ODE-Nets, and GNN-based N-body emulation. Then explicitly enumerate the paper’s novel contributions (e.g., the specific Hamiltonian + pairwise potential parameterization choice, the differentiable leapfrog unrolling setup, the radial curriculum for Plummer cores, the specific generalization tests across N) and what is inherited. Where appropriate, convert broad claims in Sec. 1 and Sec. 4 into precise, testable statements matched to experiments.

- 4. The symplecticity/phase-space preservation verification is currently underspecified and partially conceptually incorrect. Sec. 2.4 and Sec. 3.2.3 treat $\det(M) \approx 1$ as evidence/proof of a canonical/symplectic map, but determinant-one only tests volume preservation and is not sufficient for symplecticity; additionally, computing det of a $6N \times 6N$ Jacobian is numerically delicate and method-dependent (Sec. 2.4; Sec. 3.2.3).**

Recommendation: First, correct the claim: $\det(M) \approx 1$ supports volume preservation, not symplecticity. Second, clarify in Sec. 2.4 exactly how M is computed (full Jacobian via autograd vs finite differences; which N ; which timesteps; how many samples; numerical stabilization; whether $\log |\det(M)|$ is used). Third, if you want to empirically test symplecticity, report a symplectic-condition diagnostic such as $|M^T J M - J|$ (with the appropriate J) on sampled states, alongside volume diagnostics. Finally, note that (given exact gradients of a separable learned Hamiltonian) leapfrog is symplectic by construction; frame the diagnostic as an implementation/finite-precision sanity check rather than a proof.

- 5. The scope of the “generalization” claim is overstated relative to experiments. Transfer is shown only from $N = 50$ to $N = 25$ and $N = 100$ within the same data-generation family (virialized Plummer spheres, equal masses, fixed softening, fixed normalization choices). This is a narrow distribution shift and does not yet justify language like “learned a generalizable physical law” (Abstract; Sec. 3.1; Sec. 4).**

Recommendation: Narrow the claim in the Abstract/Sec. 4 to “generalizes across particle count within the family of virialized softened Plummer spheres (equal-mass, fixed ϵ).” If feasible, add tests along additional axes: (a) wider N range (including larger N), (b) different Plummer scale radii b / density scales, (c) different softening ϵ , and/or (d) different initial virial ratios. Report how rollout error and invariant metrics scale with these shifts.

- 6. Method specification is incomplete, preventing reproducibility and capacity assessment. Key architectural and training details are missing (MLP depth/width, parameter count, any normalization/residuals, optimizer hyperparameters, batch size, number of updates/epochs, window sam-**

pling/stride, random seeds). Loss details are also underspecified (what variables in MSE; whether intermediate steps contribute; exact Hamiltonian regularizer definition) (Sec. 2.2–2.3).

Recommendation: Expand Sec. 2.2 to fully specify the network (layers, hidden dims, activations, output scaling, aggregation, parameter count). Expand Sec. 2.3 to provide full training protocol (Adam hyperparameters, LR schedule, batch size, number of steps/epochs, gradient clipping, seed count, window sampling). Provide the explicit full loss formula, including whether loss is applied at only t_{n+50} or across the rollout, and whether it includes q only or (q, p) (and their relative weighting). An appendix with pseudocode/config tables would be sufficient.

7. **Energy regularization and curriculum masking may materially affect the results, but their roles are not isolated. The Hamiltonian regularizer coefficient $\lambda = 0.001$ is not justified, and without ablations it is unclear whether energy behavior/reversibility are driven by symplectic integration, the conservative force construction, the explicit energy penalty, or the curriculum (Sec. 2.3; Sec. 3.2.1–3.2.2).**

Recommendation: Add ablations in Sec. 3: train/evaluate with (i) $\lambda = 0$, (ii) a small grid of λ values, and (iii) no curriculum vs curriculum (and potentially different mask radii). Report impacts on trajectory MSE, $|\Delta H|$ (clearly defined), and reversibility error. If compute is limited, at least provide one controlled ablation per component and discuss the outcome in Sec. 3.2.

Minor issues

1. Figure 1 has a caption/graphic mismatch and unclear metric definition/normalization across N (per-particle? per-coordinate? averaged over time?); it also lacks uncertainty visualization (Fig. 1; Sec. 3.1).

Recommendation: Either (a) change Fig. 1 to time-series curves (MSE vs step/time) consistent with the text, or (b) revise caption/title/text to describe the bar-aggregated metric precisely. In the caption, define the MSE computation and normalization, and add error bars (mean \pm std or CI) with the number of test runs and seeds.

2. Figure 2 has labeling/definition ambiguity: the plot shows $|\Delta H(t)|$ but negative values appear (impossible if absolute value), and it is unclear whether ΔH is computed using the true Hamiltonian, the learned Hamiltonian, or the simulator energy. Units/normalization are also not specified (Fig. 2; Sec. 3.2.1).

Recommendation: Fix the axis label or the plotted quantity so signs/absolute value are consistent. State explicitly which Hamiltonian/energy is used for ΔH (ideally plot both “true” and “learned” energy deviations). Use a dimensionless normalization such as $\Delta H/|H(0)|$ and add run parameters (N , dt , steps) to the caption.

3. The evaluation horizon (50 steps; $T = 5.0$) is short relative to the paper’s emphasis on long-term stability in chaotic dynamics (Sec. 2.3; Sec. 3.1–3.2).

Recommendation: Add longer-horizon rollouts (e.g., $10\times$ – $100\times$ the current horizon and/or multiple dynamical times) and report whether energy error remains bounded and whether macroscopic statistics (e.g., virial ratio, radial density profile) remain stable. If not feasible, state this limitation explicitly in Sec. 4 and avoid “long-term” claims beyond the tested horizon.

4. The data-generation details are incomplete, limiting reproducibility and interpretability of energies: sampling algorithm/reference for Plummer initial conditions, normalization choices (G , total mass), treatment of center-of-mass position/velocity, and any input normalization to the network are not fully specified (Sec. 2.1).

Recommendation: In Sec. 2.1, specify the exact procedure (with citation) for sampling Plummer spheres and velocities, the target virial ratio and how it is enforced, values/conventions for G and masses, and whether data are centered to the COM frame. State any scaling/normalization applied to q and p before passing them to the network.

5. The manuscript states augmentation ensures “Galilean invariance,” but only rotations/translations are described; Galilean invariance also includes invariance under constant-velocity boosts, and COM motion handling is unclear (Sec. 2.1).

Recommendation: Either (a) rephrase to “rotational and translational invariance” and describe COM preprocessing, or (b) add explicit velocity-boost augmentation / COM velocity removal and describe it precisely in Sec. 2.1.

6. The potential form $U(q) = \frac{1}{N} \sum_{i<j} \phi(d_{ij})$ raises a scaling ambiguity across particle counts: depending on whether total mass or per-particle mass is held fixed, gravitational energy/forces scale differently with N ; this affects interpretation of “zero-shot across N ” and physical fidelity (Sec. 2.2; discussion around N -transfer in Sec. 3.1).

Recommendation: Explicitly state the scaling regime (fixed total mass vs fixed per-particle mass) and ensure $T(p)$, $U(q)$, and the simulator use consistent conventions. If $\frac{1}{N}$ is a training normalization (not physical), state that clearly and discuss implications for transferring across N .

7. No computational cost/throughput is reported, despite emulation being partly motivated by efficiency. Given the $O(N^2)$ pairwise structure, it is unclear whether the approach is faster than direct force computation for the tested N (Sec. 1; implicit in “emulation”).

Recommendation: Report wall-clock time per step (and per rollout) for the learned emulator vs the reference leapfrog simulation on the same hardware, for $N = 25/50/100$. Briefly discuss scaling with N and whether the method’s value is speed, differentiability, learned surrogate modeling, or some combination.

Very minor issues

1. An affiliation line like “Anthropic, Gemini & OpenAI servers. Planet Earth.” (as mentioned in the unstructured review) is not appropriate for a scientific manuscript.

Recommendation: Replace with standard author affiliations/institutions or remove nonstandard/placeholder affiliation text.

2. Mass and timestep notation are inconsistent: m_i vs a single m in the drift step, and dt vs Δt across sections (Sec. 2.2–2.3).

Recommendation: Standardize notation throughout. If $m_i \equiv 1$, state it once and remove ambiguous divisions; otherwise write updates explicitly componentwise with m_i . Define $\Delta t \equiv dt$ (or use one consistently).

3. Hamilton’s equations are presented incompletely ($\dot{p} = -\partial H/\partial q$ is given, but $\dot{q} = \partial H/\partial p$ is not explicitly written though it underpins the drift step) (Sec. 2.2).

Recommendation: Add $\dot{q}_i = \partial H/\partial p_i$ (and for $T(p) = \sum p^2/(2m_i)$, $\dot{q}_i = p_i/m_i$) near Eq. (3) to complete the formulation.

4. Formatting/presentation inconsistencies: mixed markdown-like headings (e.g., hash-prefixed headings), inconsistent equation numbering/typesetting, and small figure fonts reduce readability (Sec. 3; Sec. 2.2; Figs. 1–2).

Recommendation: Clean up section heading styles, ensure key equations are displayed and numbered consistently, and export figures as vector graphics with larger fonts and colorblind-safe palettes.

Key statements and references

- • **Standard machine learning emulators for N-body dynamics that treat prediction as black-box state-to-state regression typically violate conservation laws, leading to secular energy drift and unstable, divergent trajectories over long timescales, making them unsuitable for reliable modeling of chaotic gravitational systems [1–7].**
- *Reference(s):* [1], [2], [7]
- • **Hamiltonian Neural Networks and related physics-informed architectures have demonstrated that learning a scalar Hamiltonian $H(q,p)$, from which forces are obtained via automatic differentiation, yields conservative (curl-free) force fields that better respect underlying physical laws than models that directly regress forces or next states [8, 9, 12, 13].**
- *Reference(s):* [8], [9], [13]

- • Symplectic integrators such as the leapfrog (kick–drift–kick) scheme are known to preserve the geometric structure of Hamiltonian flows, including time-reversibility, phase-space volume conservation, and bounded, oscillatory energy error, in contrast to non-symplectic integrators that typically exhibit secular energy drift [14–18].
- *Reference(s)*: [14], [15], [18]

Mathematical consistency audit

This section audits **symbolic/analytic** mathematical consistency (algebra, derivations, dimensional/unit checks, definition consistency).

Maths relevance: substantial

The paper’s core method is formulated in Hamiltonian mechanics with a learned scalar potential $U(q)$, forces derived by differentiation, and a symplectic leapfrog integrator embedded in training. The mathematics present is relatively compact (few explicit equations) but central to the claims (conservativity, symplecticity/volume preservation, time reversibility, and generalization across N). Most formulas are standard and internally consistent, but a key logical overclaim appears in equating $\det(\text{Jacobian}) = 1$ with symplecticity/canonical transformation.

Checked items

- ✓ **Hamiltonian decomposition** (Sec. 2.2, p.2-3 (text: $H(q, p) = T(p) + U(q)$))
 - **Claim:** The system Hamiltonian is separable into kinetic $T(p)$ and potential $U(q)$.
 - **Checks:** definition consistency, notation consistency
 - **Verdict:** PASS; confidence: high; impact: moderate
 - **Assumptions/inputs:** Classical Hamiltonian formulation with canonical coordinates (q, p) .
 - **Notes:** Used consistently to justify kick/drift splitting in leapfrog and to derive momentum updates from $-\nabla U$.
- ✓ **Kinetic energy definition** (Sec. 2.2, p.3 (text: $T(p) = \sum p_i^2/(2m_i)$, $m_i = 1$))
 - **Claim:** Kinetic energy is $T(p) = \sum_{i=1}^N p_i^2/(2m_i)$ with unit masses.
 - **Checks:** dimensional/units, symbol consistency
 - **Verdict:** PASS; confidence: medium; impact: minor
 - **Assumptions/inputs:** p_i denotes canonical momentum for particle i ., m_i are particle masses; set to 1.
 - **Notes:** Consistent as written, but later leapfrog drift step uses a single mass symbol m ; see separate item on notation inconsistency.
- △ **Potential energy as pairwise sum with $1/N$ scaling** (Eq. (1), Sec. 2.2, p.3)

- **Claim:** $U(\mathbf{q}) = \frac{1}{N} \sum_{i < j} \phi(d_{ij})$, with shared ϕ .
- **Checks:** algebra/structure, invariance sanity-check, scaling/normalization consistency
- **Verdict:** UNCERTAIN; confidence: medium; impact: moderate
- **Assumptions/inputs:** ϕ maps a scalar distance to a scalar pair potential contribution., Sum over $i < j$ implies symmetry and permutation invariance.
- **Notes:** Permutation invariance is fine. However, the $1/N$ factor rescales forces by $1/N$ as well (since ∇U inherits the factor), which changes the implied equations of motion when N varies unless the mass/total-mass scaling regime is specified. The paper motivates $1/N$ for stability across N , but does not state the accompanying physical scaling assumptions needed for internal physical-Hamiltonian consistency across different N .

4. ✓ **Softened distance definition** (Eq. (2), Sec. 2.2, p.3)

- **Claim:** $d_{ij} = \sqrt{|\vec{q}_i - \vec{q}_j|^2 + \epsilon^2}$.
- **Checks:** dimensional/units, algebra sanity-check, differentiability/singularity check
- **Verdict:** PASS; confidence: high; impact: minor
- **Assumptions/inputs:** q_i are position vectors in \mathbb{R}^3 ., ϵ is a fixed softening length.
- **Notes:** Dimensionally consistent; removes $r = 0$ singularity and makes $\partial d_{ij} / \partial q_i$ well-defined.

5. ✓ **Forces from Hamiltonian gradient** (Eq. (3), Sec. 2.2, p.3)

- **Claim:** $\dot{p}_i = -\partial H / \partial q_i = -\partial U / \partial q_i$, implying conservative forces.
- **Checks:** derivation logic, notation consistency
- **Verdict:** PASS; confidence: high; impact: critical
- **Assumptions/inputs:** $H(\mathbf{q}, \mathbf{p}) = T(\mathbf{p}) + U(\mathbf{q})$ with U dependent only on \mathbf{q} ., q_i denotes particle i 's coordinates; derivative should be interpreted as gradient for vector \mathbf{q}_i .
- **Notes:** Correct consequence of separability: $\partial T / \partial q_i = 0$ so \dot{p}_i depends only on U . *Minor notation imprecision: for vector \mathbf{q}_i it is $\nabla_{\mathbf{q}_i} U$, not a scalar partial.*

6. ✓ **Conservative/curl-free claim from autodiff** (Sec. 2.2, p.3 (text after Eq. (3)))

- **Claim:** Deriving forces via automatic differentiation guarantees the force field is conservative (curl-free).
- **Checks:** vector calculus sanity-check, assumption check
- **Verdict:** PASS; confidence: medium; impact: moderate
- **Assumptions/inputs:** Force is defined as $\mathbf{F}(\mathbf{q}) = -\nabla U(\mathbf{q})$ for a differentiable scalar U .

- **Notes:** Given $F = -\nabla U$, the field is a gradient field and thus conservative under standard smoothness/domain assumptions. Paper does not discuss domain topology, but with softening the singularity is removed, reducing concern.

7. ✓ **Leapfrog kick update** (Sec. 2.3, p.4 (Step 1))

- **Claim:** $p_{n+\frac{1}{2}} = p_n - (\Delta t/2)\nabla_q U(q_n)$.
- **Checks:** algebra/time-discretization sanity-check, sign consistency
- **Verdict:** PASS; confidence: high; impact: critical
- **Assumptions/inputs:** Hamiltonian is separable and U depends only on q ., Canonical equation $\dot{p} = -\nabla U$.
- **Notes:** Standard half-step momentum update consistent with $\dot{p} = -\nabla U$.

8. △ **Leapfrog drift update and mass notation** (Sec. 2.3, p.4 (Step 2))

- **Claim:** $q_{n+1} = q_n + \Delta t \cdot p_{n+\frac{1}{2}}/m$.
- **Checks:** notation consistency, dimensional/units
- **Verdict:** UNCERTAIN; confidence: high; impact: minor
- **Assumptions/inputs:** $\dot{q} = \partial H/\partial p = p/m$ (for $T = p^2/(2m)$).
- **Notes:** Dimensionally fine, but inconsistent with earlier use of per-particle masses m_i and the explicit assumption $m_i = 1$. It is unclear whether m is meant to be 1, m_i , or a vector of masses. Clarification needed.

9. ✓ **Leapfrog second kick update** (Sec. 2.3, p.4 (Step 3))

- **Claim:** $p_{n+1} = p_{n+\frac{1}{2}} - (\Delta t/2)\nabla_q U(q_{n+1})$.
- **Checks:** algebra/time-discretization sanity-check, sign consistency
- **Verdict:** PASS; confidence: high; impact: critical
- **Assumptions/inputs:** Same as first kick; force evaluated at updated position.
- **Notes:** Standard symmetric kick, completing the time-reversible leapfrog step.

10. ✓ **Time reversibility test logic** (Sec. 2.4 bullet 'Time-Reversibility', p.5; Sec. 3.2.2, p.6-7)

- **Claim:** Integrating forward with dt then backward with $-dt$ should recover the initial state for a time-reversible learned dynamics/integrator.
- **Checks:** logical implication, method-property consistency
- **Verdict:** PASS; confidence: medium; impact: moderate
- **Assumptions/inputs:** The one-step map is symmetric (self-adjoint) as leapfrog is when using the same force evaluations and step size magnitude., Deterministic dynamics with no stochastic components during integration.

- **Notes:** The leapfrog scheme presented is time-reversible in exact arithmetic. The test described matches this property conceptually.
11. ✓ **Hamiltonian deviation definition** (Sec. 2.4 bullet 'Hamiltonian Conservation', p.4-5)
- **Claim:** $\Delta H(t) = |H(q(t), p(t)) - H(q(0), p(0))|$ measures energy deviation.
 - **Checks:** definition consistency, notation consistency
 - **Verdict:** PASS; confidence: high; impact: minor
 - **Assumptions/inputs:** H is the same Hamiltonian used for integration., Absolute value denotes scalar magnitude.
 - **Notes:** Consistent definition for monitoring energy stability.
12. ✗ **Phase-space Jacobian determinant criterion** (Sec. 2.4 bullet 'Phase-Space Volume Preservation', p.5; Sec. 3.2.3, p.7)
- **Claim:** For a discrete-time Hamiltonian/symplectic map, the Jacobian $M = \partial(q_{n+1}, p_{n+1})/\partial(q_n, p_n)$ must have $\det(M) = 1$; computing $\det(M) \approx 1$ provides rigorous mathematical proof of canonical/symplectic nature.
 - **Checks:** logical implication, property equivalence check
 - **Verdict:** FAIL; confidence: high; impact: critical
 - **Assumptions/inputs:** M is the full $2dN \times 2dN$ Jacobian of the one-step map.
 - **Notes:** The paper conflates volume preservation with symplecticity/canonical transformation. $\det(M) = 1$ is a volume-preservation condition (and even then, $\det \approx 1$ numerically is only suggestive), but it does not establish the stronger symplectic/canonical condition on its own. Therefore the statement that this provides 'rigorous mathematical proof' of symplecticity/canonical transformation is not internally justified by the criterion given.
13. ✗ **Galilean invariance augmentation statement** (Sec. 2.1, p.3 (last paragraph))
- **Claim:** Random 3D rotation and translation augmentation ensures the model learns Galilean invariance.
 - **Checks:** definition consistency
 - **Verdict:** FAIL; confidence: high; impact: minor
 - **Assumptions/inputs:** Galilean invariance includes invariance under translations, rotations, and constant-velocity boosts.
 - **Notes:** Rotation+translation address Euclidean invariance but do not enforce invariance to velocity boosts. This is a definitional mismatch (math/physics symmetry statement), though it may not affect the core Hamiltonian/leapfrog mathematics.

Limitations

- Audit is restricted to the provided PDF text; no supplementary material, appendices, or code-level definitions were available to resolve ambiguities (e.g., state uses (q, v) vs (q, p) , mass scaling across N).
- Several key claims (e.g., symplecticity proof, generalization scaling) are asserted without full analytic derivations; where steps are missing, items are marked UNCERTAIN rather than filled in.
- Figures were not used to validate any mathematical properties; numerical/empirical validation is out of scope by instruction.

Numerical results audit

This section audits **numerical/empirical** consistency: reported metrics, experimental design, baseline comparisons, statistical evidence, leakage risks, and reproducibility.

Six algebraic/unit-consistency checks over explicitly stated quantities (dt , T , steps, dataset counts, horizon definition, ϵ consistency, sign conventions for reverse integration, and a curriculum threshold substitution) were executed and all passed with $diff_{abs} = 0$ and $diff_{rel} = 0$ within the stated tolerances.

Checked items

1. ✓ **C1_steps_vs_T_dt** (Sec. 2.1 (page 3): " $dt = 0.01$ "; " $T = 5.0$ " and Sec. 3.1 (page 5): "500 steps ($T = 5.0$)")
 - **Claim:** The paper states $dt = 0.01$ and total duration $T = 5.0$, and later refers to a full integration of 500 steps corresponding to $T = 5.0$.
 - **Checks:** unit-consistent recomputation ($steps = T/dt$)
 - **Verdict:** PASS
 - **Notes:** Computed steps = $T/dt = 5.0/0.01 = 500.0$.
2. ✓ **C2_train_test_split_total_sims** (Sec. 2.1 (page 3): "A total of 100 independent simulations... 80 ... training and 20 ... testing.")
 - **Claim:** The train/test split (80/20) sums to the stated total of 100 simulations.
 - **Checks:** parts-to-total sum
 - **Verdict:** PASS
 - **Notes:** Computed train+test = $80.0 + 20.0 = 100.0$.
3. ✓ **C3_horizon_steps_vs_50dt** (Sec. 2.3 (page 4): "MSE between ... at $t_n + 50 \cdot dt$ "; Sec. 2.1 (page 3): " $dt = 0.01$ ")
 - **Claim:** The prediction horizon $t_n + 50 \cdot dt$ corresponds to 0.5 time units given $dt = 0.01$.
 - **Checks:** unit-consistent recomputation (time horizon)

- **Verdict:** PASS
 - **Notes:** Computed horizon = $50.0 \times 0.01 = 0.5$.
4. ✓ **C4_softening_length_consistency** (Sec. 2.1 (page 3): " $\epsilon = 0.01$ "; Sec. 2.2 (page 3): " $\epsilon = 0.01$... same ... used in ... data generation.")
- **Claim:** The gravitational softening length epsilon is stated as 0.01 in both the simulation setup and the model architecture, and should match exactly.
 - **Checks:** repeated constant equality across sections
 - **Verdict:** PASS
 - **Notes:** Compared epsilon values across sections.
5. ✓ **C5_negative_dt_magnitude_matches_dt** (Sec. 2.1 (page 3): " $dt = 0.01$ "; Sec. 3.2.2 (page 7): " $dt = -0.01$ ")
- **Claim:** The backward integration timestep $dt = -0.01$ should have magnitude equal to the forward timestep 0.01.
 - **Checks:** sign/magnitude consistency
 - **Verdict:** PASS
 - **Notes:** Checked magnitude and sign: $|dt_{backward}|$ vs $dt_{forward}$ used for diff aggregation.
6. ✓ **C6_curriculum_threshold_numeric** (Sec. 2.3 (page 4): " $r > 0.5b$, where $b = 1$ ")
- **Claim:** Given $b = 1$, the curriculum masking threshold $r > 0.5b$ implies $r > 0.5$ in the same units as r .
 - **Checks:** algebraic substitution
 - **Verdict:** PASS
 - **Notes:** Computed threshold = $0.5 \times 1.0 = 0.5$.

Limitations

- Only parsed text from the provided PDF pages was used; figures are not machine-readable here for exact numeric extraction.
- Checks are limited to algebraic consistency among explicitly stated numbers (e.g., dt , T , counts, constants); claims based on experimental results without tabulated values cannot be recomputed.
- No external datasets, simulation outputs, model weights, or code are available in the PDF text to verify performance metrics, Jacobian determinants, or energy/reversibility error magnitudes.