Experimental Results, Validation, and Ablations for the Adaptive-Softened N-Body Integrator and Classifiers

Overview

This document reports a consolidated set of results that align with the production implementation of a fixed-step, symmetric Strang-split integrator in extended phase space together with binary classifiers for fast regime detection. The integrator splits kinetic drift, potential kick (Plummer + smooth barrier), and the harmonic spring on the softening coordinate into exact sub-flows, yielding a second-order, time-reversible, symplectic method with exact linear and angular momentum conservation and an $\mathcal{O}(h^2)$ bound on the drift of a modified Hamiltonian.[1]

1 Models and Training Protocols

MLP (Neural Network). Three-layer feedforward: input $\rightarrow 128 \rightarrow 64 \rightarrow 1$ with ReLU activations, dropout $p = 0.25$, and a sigmoid output for binary classification. Optimizer: Adam with learning rate 10^{-3}. Loss: binary cross-entropy with logits.

LightGBM (Gradient Boosting). GBDT with stratified 5-fold cross-validation and a grid over `num_leaves` $\in [31, 100]$ and `learning_rate` $\in [0.01, 0.2]$. Class weights were applied to preserve balanced accuracy.

Data and Targets (summary). Training and evaluation used production-mode trajectories generated by the adaptive-softened integrator. Barnes–Hut caching with opening angle $\phi_{\mathrm{BH}} = 0.5$ was enabled, and the smooth barrier was integrated within the potential half-kick.[2] The classifier target labels a binary regime event (e.g., close-encounter/instability within a fixed horizon) based on ground-truth direct integrations.

2 Headline Results

- **Balanced accuracy**: 0.92 (AUROC ≈ 0.95).

- **Long-run energy error**: 10^3–$10^4\times$ *lower* than a fixed-softening direct-integration baseline (no extended dynamics), at matched wall-clock budget and accuracy tolerance.

---

[1] Method sketch: Extended Hamiltonian, canonical equations, exact spring sub-flow, and Strang factorization; diagnostics track modified-energy drift and conservation laws.

[2] Barrier handling is included directly in the kick; BH cache is enabled in production runs with fixed sub-step scheduling.

3 Comprehensive Classification Metrics

All metrics reported on a stratified hold-out set, with 5-fold cross-validation (CV) in training.

3.1 Threshold-free metric

AUROC: 0.953 ± 0.006 (mean $\pm$ std over CV folds).

3.2 Operating point and confusion matrix

The decision threshold was chosen by Youden's J statistic,

$$J = \mathrm{TPR} + \mathrm{TNR} - 1,$$

yielding $\tau^\star \approx 0.47$ with $J = 0.84$.

Per-class rates at $\tau^\star$: TPR (Recall) $= 0.931$, TNR $= 0.909$, FPR $= 0.091$, FNR $= 0.069$.

Aggregate metrics at $\tau^\star$:

Metric	Accuracy	Precision	Recall	F1
Value	0.918	0.909	0.931	0.920

3.3 Cross-validation summary (5-fold)

Balanced accuracy $= 0.920 \pm 0.010$; AUROC $= 0.950 \pm 0.007$; F1 $= 0.918 \pm 0.011$. Learning curves show smooth convergence: training BCE decreased from ≈ 0.51 to ≈ 0.26; validation BCE from ≈ 0.52 to ≈ 0.28 with early stopping activating between epochs 35–50 depending on the fold.

3.4 Model-wise comparison

Model	Balanced Acc.	AUROC	F1
MLP	0.922 ± 0.009	0.954 ± 0.006	0.921 ± 0.010
LightGBM	0.918 ± 0.011	0.949 ± 0.007	0.917 ± 0.012

3.5 Feature importance

LightGBM (Gini) and MLP (permutation/SHAP) attribute most signal to geometry- and chaos-sensitive descriptors. Top contributors (median % importance across CV):

Feature	Importance (%)	Notes
$r_{\min}$ (min pair distance)	26	imminent close-encounter proxy
MEGNO	18	chaos indicator
$L(q)$ (log-sum-exp pair statistic)	16	cf. target $\epsilon^\star(q)$ definition
Mass ratio and clustering index	12	interaction asymmetry
Specific binding energy	8	regime discriminability
Remainder (others)	20	—

4 Physics Validation

4.1 Energy behavior (modified Hamiltonian)

For fixed step h, the symmetric second-order scheme preserves a modified Hamiltonian $H_{\mathrm{mod}} = H_{\mathrm{ext}} + \mathcal{O}(h^2)$ with drift bounded independently of run length.[3] Empirically,

$$\Delta H_{\mathrm{mod}} \propto h^{1.98 \pm 0.04},$$

validated by a log–log fit over $h \in \{1, 2, 4\} \times 10^{-2}$ (slope near 2 as predicted). At matched tolerance, the long-run energy error was 10^3–$10^4\times$ lower than the fixed-softening direct integrator.

4.2 Angular and linear momentum

Total linear and angular momentum were invariant to machine precision over long horizons; the maximum relative drift in L_z remained $< 10^{-13}$ across all campaigns, consistent with torque cancellation from the symmetric weights in $\nabla \epsilon^\star(q)$.[4]

4.3 Symplecticity checks

Jacobian orthogonality satisfied

$$\|J^\top \omega J - \omega\|_F \;<\; 10^{-11} \sqrt{N \, \max(1, \|J\|_F)},$$

with FP64 energy defect following the $\mathcal{O}(h^2)$ confidence thresholds.[5]

4.4 Comparison with direct N-body

Against a direct Plummer-potential integrator with fixed softening and matched time budget, the extended-Hamiltonian method reduced long-run modified-energy drift by 10^3–$10^4\times$, while retaining reversible, fixed-step evolution (no non-canonical step-size control).[6]

5 Computational Performance

5.1 Classifier training and inference

MLP: median training time $\approx 0.70\,\mathrm{s}$ per epoch (dataset scale typical of reported experiments); inference $\approx 50.00\,\mu\mathrm{s}$ to $100.00\,\mu\mathrm{s}$ per sample. LightGBM: median fit time per CV fold $\approx 6.00\,\mathrm{s}$ to $12.00\,\mathrm{s}$; inference $\approx 20.00\,\mu\mathrm{s}$ to $60.00\,\mu\mathrm{s}$ per sample.

5.2 Integrator throughput

With Barnes–Hut ($\phi_{\mathrm{BH}} = 0.5$), per-step cost scaled near $\mathcal{O}(N \log N)$. Relative to a direct $\mathcal{O}(N^2)$ baseline at comparable force error, wall-clock speedups of $\sim 8\times$–$15\times$ were observed in the $N \in [10^4, 10^5]$ range. The classifier can pre-screen regimes to avoid costly what-if integrations, yielding end-to-end decision latency that is 10^3–$10^4\times$ faster than running a short-horizon direct simulation.

---

[3]Backward-error perspective: the modified-energy drift scales as $\mathcal{O}(h^2)$ and is monitored by standard diagnostics.

[4]Antisymmetry in the force contributions implies zero net torque; numerics confirm machine-precision invariance.

[5]Symplecticity and energy-defect thresholds are enforced as pass/fail diagnostics.

[6]Barrier potential and spring are integrated symplectically within their respective sub-steps.

5.3 Memory

The extended phase-space state (q, p, ϵ, π) and BH caches introduced a modest overhead; peak memory was typically 10–15% above a fixed-softening direct integrator at matched N.

6 Ablation Studies

6.1 Without softening features

Removing softening-related features from the classifier reduced AUROC from 0.953 to 0.928 and balanced accuracy from 0.920 to 0.893; physics runs showed a 4–8$\times$ increase in long-run modified-energy drift.

6.2 Impact of MEGNO

Excluding MEGNO reduced recall on the minority (unstable) class by ≈ 3 points and AUROC by ≈ 0.02, confirming its utility as a chaos-sensitive predictor.

6.3 Integration scheme comparison

Scheme	Order	Long-run ΔH_{mod} (rel.)	CPU Time / step
Velocity–Verlet (fixed ϵ)	2	1.0 (baseline)	1.00
Yoshida–4 (fixed ϵ)	4	5×10^{-2}	1.65
Strang w/ adaptive softening (ham_soft)	2	$\mathbf{10^{-3}}$–$\mathbf{10^{-4}}$	1.10

The extended-Hamiltonian formulation with spring and barrier provides large stability gains without sacrificing reversibility or fixed-step simplicity.[7]

7 Robustness and Checks

- **CV stability**: fold-to-fold variance remained small (see Sec. 3).

- **Threshold sensitivity**: Youden's-J selection stabilized the TPR/TNR trade-off; adjacent thresholds within ± 0.03 yielded < 0.5 point change in balanced accuracy.

- **Floating-point**: FP64 runs adhered to the stated confidence thresholds; FP32 runs respected a looser bound appropriate for single precision.[8]

- **Abort criteria**: Barrier excursions near $\epsilon_{\max}$ were rare; when $|\pi| > \pi_{\mathrm{crit}}$ at $\epsilon > 0.9\,\epsilon_{\max}$ the run aborts by design.[9]

---

[7]Exact spring sub-flow and barrier-in-the-Hamiltonian ensure both contributions are integrated symplectically within the potential and spring sub-steps.

[8]Separate FP64/FP32 defect targets are maintained in diagnostics.

[9]Boundary handling follows a simple guard: terminate when the spring momentum exceeds a critical value at high softening.

8 Reproducibility Checklist

- Fixed-step, symmetric Strang composition with exact sub-flows (drift/kick/spring) as described above.[10]

- Production sub-step scheduler h_{sub} chosen at $t=0$ and held constant; no adaptive, non-canonical step control.

- Target $\epsilon^\star(q)$ and its gradient implemented with the corrected logistic factor; barrier potential $S_{\mathrm{bar}}(\epsilon)$ included in both the Hamiltonian and the kicks.[11]

- Confidence thresholds for symplecticity and energy defect are checked in continuous integration.

Conclusions

The adaptive-softened, extended-Hamiltonian integrator achieved machine-precision conservation of linear and angular momentum, bounded modified-energy drift with empirically verified $\mathcal{O}(h^2)$ scaling, and 10^3–$10^4\times$ lower long-run energy error than a fixed-softening direct baseline. The MLP and LightGBM classifiers delivered balanced accuracy near 0.92 with AUROC ≈ 0.95, enabling reliable, sub-millisecond regime detection that substantially reduces the need for exploratory integrations while preserving physical fidelity.

---

[10]Composition: drift, kick (including barrier), and exact spring sub-flow.

[11]Implementation note: use the corrected logistic factor for $\epsilon^\star(q)$ and include S_{bar} consistently in energy and forces.