Symplectic Parallel Scan: A Neural Hamiltonian Framework for Accelerated Scientific Simulation

Sungwoo Park
Korea University
sungwoo_park@korea.ac.kr

Abstract

Neural Hamiltonian models provide a principled approach to learning dynamical systems by embedding the symplectic structure of physical evolution into the model class. However, this geometric prior alone has not made Hamiltonian neural networks efficient long-horizon simulators: an M-step trajectory is still typically generated by M sequential updates, and common training objectives based on force matching or one-step prediction provide only local supervision. We introduce *Symplectic Parallel Scan (SPS)*, a neural Hamiltonian framework that algebraizes the learned dynamics through a finite-dimensional Poisson algebra of Hamiltonian generators. SPS constructs a Neural–Poisson Composition DAG closed under projected products and Poisson brackets, whose elements induce symplectic Hamiltonian drivers organized as a Lie-group walk. The associativity of this flow composition enables a parallel prefix-scan over Hamiltonian drivers, reducing the sequential depth of trajectory propagation from $\mathcal{O}(M)$ to $\mathcal{O}(\log M)$ while keeping each prefix inside the symplectic flow family. The resulting model is trained through a global trajectory-matching objective rather than only local force labels. Across quantum spin and molecular dynamics benchmarks, SPS achieves substantial wall-clock acceleration over sequential HNN baselines, improves long-horizon prediction accuracy, and reduces energy drift and symplectic violation under extrapolation.

1 Introduction

Learning data-driven models that respect the *symplectic* structure of Hamiltonian mechanics has emerged as a principled and rapidly evolving direction within the broader AI4Science initiative, driven by the need for robust long-horizon prediction in physics [Li et al., 2020], control, and molecular dynamics simulation [Zhang et al., 2025a]. Symplectic neural integrators and Hamiltonian neural networks (HNNs) have shown that embedding canonical invariants into the hypothesis space can mitigate energy drift and stabilize learning over moderate horizons. These developments have established neural Hamiltonians as a promising substrate for *scientific simulation*, where the goal is not to interpolate observed trajectories but to act as surrogates for high-fidelity numerical solvers.

Despite this progress, existing neural Hamiltonian models are fundamentally misaligned with the regimes that matter most for AI-based accelerated scientific simulation. Firstly, conventional simulation methods such as HNNs advance time by composing a learned symplectic update M times to reach a horizon of M steps, yielding an inherently sequential $\mathcal{O}(M)$ critical path. On modern accelerators, this linear depth quickly becomes the dominant bottleneck in long-horizon rollouts, even when each step is cheap and highly optimized. Moreover, this sequential dependence is *compounded* by the nonlinear black-box structure of the learned force field: the update at step $k+1$ depends on the full state produced at step k, so standard acceleration strategies such as operator splitting and multi-rate integrators [Bartel et al., 2023] are inapplicable without breaking the underlying symplectic

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

structure. In practice, this means that current symplectic-learning methods cannot serve as genuinely *accelerated* surrogates for large-scale Hamiltonian workloads where wall-clock time is critical.

Secondly, prevailing HNN frameworks are typically trained under a *point-wise force-matching objective*. An unconstrained generator is optimized to regress the instantaneous accelerations $(\dot{\mathbf{q}}, \dot{\mathbf{p}}) = (\nabla_p H_\theta, -\nabla_q H_\theta)$ at each observed state in phase space. This local criterion neglects that many scientific settings pose a *global inverse problem*: only initial conditions and chaotic trajectory traces are observed, while the latent Hamiltonian (*e.g.*, force field) must be inferred so as to reproduce the entire path. As a result, models calibrated by instantaneous force regression often extrapolate poorly. They remain empirically stable for a few characteristic time scales but drift off physically admissible solution manifolds when driven far beyond the training distribution, precisely in the long-time regime where accelerated simulators are most needed.

In parallel, recent advances in *parallel scan (prefix-sum)* algorithms [Hillis and Steele Jr., 1986, Blelloch, 1990] within the ML systems community offer a compelling route toward breaking the temporal bottleneck. By reorganizing computations into associative binary trees, scans reduce the critical depth of a product of M operators from $\mathcal{O}(M)$ to $\mathcal{O}(\log M)$, enabling massive GPU/TPU throughput. This strategy has already underpinned large-scale attention [Dao et al., 2022] and structured state-space sequence models [Gu et al., 2022, Smith et al., 2023]. However, conventional Hamiltonian generators produce nonlinear flow maps whose composition is not expressed in terms of a simple associative algebra of parameters. Thus, there is no obvious way to rewrite a learned symplectic integrator as an associative binary operation for scan-based acceleration without sacrificing Hamiltonian structure.

In this work, we close this gap by explicitly embedding neural Hamiltonians into a carefully constructed *symplectic* framework tailored for accelerated scientific simulation. We parameterize the Hamiltonian as a structured polynomial built from primitive blocks that are closed under the Poisson bracket [Marsden and Ratiu, 1999, Abraham and Marsden, 1978]. This Poisson-closed dictionary induces a finite-dimensional Lie algebra of Hamiltonian vector fields whose exponentials form a Lie group of symplectic flow maps, *i.e.*, *symplectomorphism* [McDuff and Salamon, 1998, Arnold, 1989]. At the same time, the composition of these flows is governed by an *associative* algebra of coefficients, enabling long-horizon evolution to be realized as a parallel prefix product on this Lie group. Thus, the proposed construction addresses the two central shortcomings above: it provides a log-depth, structure-preserving neural integrator suitable for high-throughput accelerators, and it encourages globally consistent dynamics beyond the training horizon. Overall, our contributions in this work are therefore twofold:

(1) Acceleration of Physical Simulation with Parallel Scan. We introduce the *Symplectic Parallel Scan*, a principled acceleration methods for neural Hamiltonians by leveraging Poisson algebraic decomposition. The proposed method achieves symplecticity and up to $10^3\times$ **acceleration** compared to existing Hamiltonian Neural Network (HNN) variants, by enabling logarithmic-time trajectory prediction through an associative prefix-scan approach.

(2) Symplectic Structured Inverse Problem. We establish that our trajectory-matching framework robustly solves the inverse Hamiltonian learning problem. By leveraging global trajectory consistency rather than local force labels, our method provides stable and physically meaningful extrapolation beyond the training regime, significantly outperforming conventional force-matching approaches that often suffer from instability during extrapolation.

To lay the groundwork for our framework, we first briefly review the necessary mathematical preliminaries and notation. Let $(\mathbb{R}^{2n}, \omega)$ be a $2n$-dimensional symplectic space, and let $z := (\mathbf{q}, \mathbf{p}) = (q_1, \ldots, q_n, p_1, \ldots, p_n)$ denote canonical coordinates associated with the standard Darboux form $\omega = \sum_{i=1}^n dq_i \wedge dp_i$. For each sufficiently smooth scalar function $g \in C^\infty(\mathbb{R}^{2n})$, referred to as a *Hamiltonian generator*, we associate a unique vector field X_g, called the *Hamiltonian vector field* of g, implicitly defined through the relation $\iota_{X_g}\omega = dg$, explicitly yielding $X_g = \sum_{i=1}^n (\partial_{p_i} g)\partial_{q_i} - (\partial_{q_i} g)\partial_{p_i}$. Furthermore, given any two smooth functions $f, g \in C^\infty(\mathbb{R}^{2n})$, the *Poisson bracket* is defined as $\{f, g\} := \omega(X_f, X_g) = \sum_{i=1}^n (\partial_{q_i} f \partial_{p_i} g - \partial_{p_i} f \partial_{q_i} g)$, providing a fundamental algebraic structure for analyzing Hamiltonian dynamics.

Problem Formulation. Recall that the primary goal of *Hamiltonian neural networks* (HNNs) is to approximate an unknown scalar-valued Hamiltonian function $H_{\text{true}} : \mathbb{R}^{2n} \to \mathbb{R}$ using a neural network $g_\theta : \mathbb{R}^{2n} \to \mathbb{R}$. Upon successful training, the learned model g_θ serves as a Hamiltonian

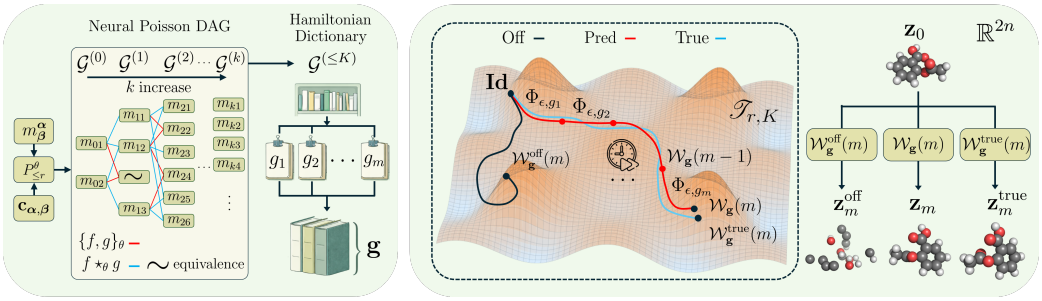


Figure 1: **Pipeline of Proposed Method.** *(left) Construction of an ordered Hamiltonian dictionary.* Algorithm 1 predicts neural coefficients to instantiate the polynomial basis, and Algorithm 2 constructs a Neural–Poisson DAG, producing an ordered Hamiltonian dictionary $\mathbf{g}$. *(right) Phase-Space state evolution from Lie–Group walk.* We show a Lie-group walk defined by successive composition of symplectic flows $\Phi_{\epsilon,g}$, with accelerated computation via a parallel prefix scan, comparing an energy-drifting walk $\mathcal{W}_{\mathbf{g}}^{\text{off}}(m)$, the predicted walk $\mathcal{W}_{\mathbf{g}}(m)$, and the true walk $\mathcal{W}_{\mathbf{g}}^{\text{true}}(m)$ from $\mathbf{Id} \in \mathscr{T}_{r,K}$. The corresponding phase-space evolution is given by $z_m = \mathcal{W}_{\mathbf{g}}(m) z_0 \in \mathbb{R}^{2n}$.

generator, governing the dynamical evolution via the Poisson bracket $\{\cdot,\cdot\}$:

$$\dot{z} = \underbrace{\{z, g_\theta\} = J\nabla_z g_\theta(z)}_{\text{Learned Dynamics}} \approx \underbrace{J\nabla_z H_{\text{true}}(z) = \{z, H_{\text{true}}\}}_{\text{True Dynamics}}, \quad J = \begin{bmatrix} 0 & I_n \\ -I_n & 0 \end{bmatrix}, \tag{1}$$

where J denotes the symplectic matrix and $z \in \mathbb{R}^{2n}$ represents the state vector in physical space. To achieve this, we consider an initial condition $\hat{z}_0 \sim \mathbb{P}_{\text{init}}$ with corresponding trajectory data $\hat{z}_t \sim \mathbb{P}_{t>0}$. This setting defines a *physical inverse problem*, aiming to identify the parametrized Hamiltonian $g_\theta(z)$ by solving the following initial value optimization problem:

$$\min_{g_\theta \in \mathcal{H}} \mathbb{E}_{\hat{z}_0 \sim P_{\text{init}}, t \sim p_t} \left[\mathbb{E}_{\hat{z}_t \sim P_{t>0}} \left\| z_t(g_\theta) - \hat{z}_t \right\|^2 \middle| \hat{z}_0 \right] \xrightarrow[\text{Section 2}]{\textbf{Algebraization}} \mathcal{H} \ni g_\theta \approx H_{\text{true}}. \tag{2}$$

In our framework, we denote by $\mathcal{H}$ the hypothesis space in which the Hamiltonian generator is learned. Rather than allowing g_θ to range over arbitrary neural function classes, we propose to construct the hypothesis space $\mathcal{H}$ of Hamiltonian generators through an *algebraization* procedure. Section 2 develops this construction explicitly by expanding the generator in a Poisson-algebraic basis.

2 Model: Algebraic Hypothesis Space for Hamiltonian Dynamics

Before introducing the core machinery, we first delineate the *hypothesis space* $\mathcal{H}$ on which our model is built, ensuring that all subsequent constructions rest on a clear and consistent foundation.

Algebraic Hypothesis Space. We begin by defining a structured hypothesis class $\mathcal{H}$ for the Hamiltonian generator. Instead of using an unconstrained black-box neural network, we aim to represent g_θ as a finite algebraic expansion with neural network-predicted coefficients $\mathbf{c}_{\boldsymbol{\alpha},\boldsymbol{\beta}}$, which we call a *neural Poisson-algebraic generators*. This representation preserves neural expressivity through the coefficient map while exposing an explicit algebraic structure:

Definition 2.1 (Neural Poisson-Algebraic Generator)**.** Let us consider neural network $\mathbf{c}_{\boldsymbol{\alpha},\boldsymbol{\beta}} := c(\boldsymbol{\alpha}, \boldsymbol{\beta}, z_0, t; \theta) \in \mathbb{R}^d$, called *neural coefficients*. Given Darboux coordinates $(\mathbf{q}, \mathbf{p})$ with monomials $m_{\beta^k}^{\alpha^k}(z)$ of degree $k \le r$, we define the associated family of parameterized function as

$$g_\theta(q,p) \in P_{\le r}^\theta := \left\{ \sum_{k=1}^{|\mathcal{C}|} \mathbf{c}_{\boldsymbol{\alpha},\boldsymbol{\beta}} \cdot m_{\boldsymbol{\beta}}^{\boldsymbol{\alpha}} \;\middle|\; |\mathbf{c}_{\boldsymbol{\alpha},\boldsymbol{\beta}}| = \binom{2n+r}{r} \right\}. \tag{3}$$

In $2n$-dimensional phase spaces with coordinates $z = (q_1, \ldots, q_n, p_1, \ldots, p_n)$ and polynomial multi-indices $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{N}^n$, we represent the neural polynomial Hamiltonian generator explicitly as:

$$\textbf{(Neural Hamiltonian Generator)} \quad g_\theta(q,p) = \sum_{\boldsymbol{\alpha},\boldsymbol{\beta}} \underbrace{c(\boldsymbol{\alpha}, \boldsymbol{\beta}, z_0, t; \theta)}_{=\mathbf{c}_{\boldsymbol{\alpha},\boldsymbol{\beta}}} \cdot \underbrace{q_1^{\alpha_1} \cdots q_n^{\alpha_n} p_1^{\beta_1} \cdots p_n^{\beta_n}}_{m_{\boldsymbol{\beta}}^{\boldsymbol{\alpha}} = q_n^{\boldsymbol{\alpha}} p_n^{\boldsymbol{\beta}}}. \tag{4}$$

Table 1: **Example construction of a Neural–Poisson Composition DAG for** $(r, n) = (2, 2)$. Starting from the neural polynomial basis $\mathcal{G}^{(0)} = P^{\theta}_{\leq r}$, higher-depth nodes are generated by truncated products and Poisson brackets, while algebraically redundant terms are removed via product equivalence and the Jacobi identity.

$\mathcal{G}^{(0)}$: $\Big\{c_1q_1 + c_2q_2 + c_3p_1 + c_4q_1q_2 + c_5q_1p_2 + c_6q_2p_1 + c_7q_2p_2 + +c_8p_1p_2 + c_9p_2^2, \dots\Big\} \coloneqq P^{\theta}_{\leq r}$ **in Def 2.1**

$\mathcal{G}^{(1)}$: $\Big\{c_{12}q_1 \star_\theta q_2, \cancel{c_{21}q_2 \star_\theta q_1}, c_{23}\{q_2, p_1\}, c_{24}q_2^2 \star_\theta p_2, c_{34}\{q_1^2, q_1q_2\}, \dots\Big\}$, **Equivalence in Product**

$\mathcal{G}^{(2)}$: $\Big\{\cancel{c_{123}\{\{q_1, q_2\}, p_1\}}, c_{231}\{\{q_2, p_1\}, q_1\}, c_{312}\{\{p_1, q_1\}, p_2\}, \dots\Big\}$, **Equivalence in Jacobi Identity**

Having the definition in one hand, the original Hamiltonian dynamics introduced in (1) can be explicitly rewritten by using the learned generator as follows:

$$\textbf{(Depth 0)} \quad \dot{z} = J\nabla_z g^{(0)}_\theta(z) = \left[\dot{q}_i = \sum_{\boldsymbol{\alpha},\boldsymbol{\beta}} c^{(0)}_{\boldsymbol{\alpha},\boldsymbol{\beta}} \beta_i q^{\boldsymbol{\alpha}} p^{\boldsymbol{\beta}-\mathbf{e}_i}, \;\; \dot{p}_i = -\sum_{\boldsymbol{\alpha},\boldsymbol{\beta}} c^{(0)}_{\boldsymbol{\alpha},\boldsymbol{\beta}} \alpha_i q^{\boldsymbol{\alpha}-\mathbf{e}_i} p^{\boldsymbol{\beta}}\right]^\top.$$

While the Poisson construction above gives an explicit Hamiltonian dynamics, the direct neural polynomial generator remains restrictive when used in isolation: even though the coefficients $c_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ are produced by a neural network, the dependence on the phase-space variables (q, p) is confined to the fixed monomial dictionary $q^{\boldsymbol{\alpha}}p^{\boldsymbol{\beta}}$ and its canonical derivatives, and leading limited representability.

Neural-Poisson Composition DAG. To enrich the representation in (4) without leaving the symplectic structure (*e.g.*, preservation of the symplectic form and phase-space volume), we aim to exploit the algebraic properties of Hamiltonian dynamics. Specifically, we first view the hypothesis space $\mathcal{H}$ in (3) as a *Poisson algebra* [Marsden and Ratiu, 1999] $\big(P^{\theta}_{\leq r}, \star_\theta, \{\cdot,\cdot\}_\theta\big)$, where for any $f, g \in P^{\theta}_{\leq r}$, the dual operations are defined by

$$\pi_{\leq r}\left(\sum_{\boldsymbol{\alpha},\boldsymbol{\beta}} c_{\boldsymbol{\alpha},\boldsymbol{\beta}} q^{\boldsymbol{\alpha}} p^{\boldsymbol{\beta}}\right) \coloneqq \sum_{|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|\leq r} c_{\boldsymbol{\alpha},\boldsymbol{\beta}} q^{\boldsymbol{\alpha}} p^{\boldsymbol{\beta}}, \quad f \star_\theta g \coloneqq \pi_{\leq r}(fg), \quad \{f, g\}_\theta \coloneqq \pi_{\leq r}\big(\{f, g\}\big).$$

The projection $\pi_{\leq r}$ defines the finite Poisson-algebraic universe in which all subsequent operations are performed. Based on the introduced Poisson algebra, our aim is to recursively build a *directed acyclic graph* (DAG) of Hamiltonian generators for enlarging the q- and p-dependency beyond the limited monomial representability of (4). The base 0-th depth layer $\mathcal{G}^{(0)}$ is set to the constructed $P^{\theta}_{\leq r}$ in (3), and each subsequent layer is obtained by composing existing generators through $\star_\theta$ and $\{\cdot,\cdot\}_\theta$:

Proposition 2.2 (Neural–Poisson Composition DAG)**.** *For each $k \geq 1$ with $f^{(k-1)}, g^{(k-1)} \in \mathcal{G}^{(k-1)}$ and $\mathcal{G}^{(0)} = P^{\theta}_{\leq r}$, the following full composition graph $\mathcal{G}^{(\leq K)} \coloneqq \bigcup_{k=0}^{K} \mathcal{G}^{(k)}$ composes a DAG of depth K of Poisson Hamiltonian generators:*

$$\mathcal{H} \coloneqq \mathcal{G}^{(k)} = \left\{ g^{(k)}_\theta \,\middle|\, \{\pi_{\leq r}(f^{(k-1)} \star_\theta g^{(k-1)})\} \bigcup \{\pi_{\leq r}(\{f^{(k-1)}, g^{(k-1)}\}_\theta)\} \,/\sim \right\},$$

where equivalence relation $\sim$ induces quotient space with following three:

$$f \star_\theta g \sim g \star_\theta f, \; \{f, g\}_\theta \sim -\{g, f\}_\theta, \; \{f, \{g, h\}_\theta\}_\theta + \{g, \{h, f\}_\theta\}_\theta + \{h, \{f, g\}_\theta\}_\theta \sim 0.$$

The key consequence of this construction is that every newly generated node remains inside the Poisson-algebraic hypothesis space $\mathcal{H}$. Since the operations $\star_\theta$ and $\{\cdot,\cdot\}_\theta$ are closed after the degree-truncation projection, each element still defines a valid Hamiltonian generator and therefore induces a Hamiltonian vector field $X_g = J\nabla_z g$. Specifically, Theorem A.2 formalizes this property by showing that the dynamics generated by the Neural-Poisson Composition DAG preserve the *symplectic structure* and phase-space volume inherited from Hamiltonian mechanics, and preserve the target energy in the ideal zero-loss regime of training. Within the k-th note $\mathcal{G}^{(k)}$ of DAG, the Hamiltonian generator $g^{(k)} \in \mathcal{G}^{(k)}$ explicitly defines the corresponding Hamiltonian dynamics:

$$\textbf{(Depth } k\textbf{)} \quad \dot{z} = J\nabla_z g^{(k)}_\theta(z) = \left[\sum_{\boldsymbol{\alpha},\boldsymbol{\beta}} c^{(k)}_{\boldsymbol{\alpha},\boldsymbol{\beta}} \boldsymbol{\beta} q^{\boldsymbol{\alpha}} p^{\boldsymbol{\beta}-\mathbf{e}}, -\sum_{\boldsymbol{\alpha},\boldsymbol{\beta}} c^{(k)}_{\boldsymbol{\alpha},\boldsymbol{\beta}} \boldsymbol{\alpha} q^{\boldsymbol{\alpha}-\mathbf{e}} p^{\boldsymbol{\beta}}\right]^\top, \quad g^{(k)}_\theta \in \mathcal{G}^{(k)}. \tag{5}$$

Table 1 illustrates the example of DAG construction, showing that the quotient relation $\sim$ removes algebraically redundant DAG nodes generated by *commutativity*, *skew-symmetry*, and the *Jacobi identity*. The quotient therefore yields a compact set of representative Hamiltonian generators without changing their span. At the same time, the recursive product and Poisson-bracket operations enrich the hypothesis space $\mathcal{H}$ by forming higher-order combinations of the q- and p-orders beyond the original space, *i.e.*, $P^{\theta}_{\leq r} := \mathcal{G}^{(0)} \subseteq \cdots \subseteq \mathcal{G}^{(k)} \subset \cdots \subseteq \mathcal{G}^{(K)} := \mathcal{H}$, as illustrated in Fig 1*(left)*.

3 Acceleration : Poisson-Algebra Meets Parallel Scan

In the previous section, we constructed the Neural–Poisson Composition DAG, which enlarges the Hamiltonian-generator hypothesis space while preserving its Poisson-algebraic structure. The primary goal of this section is to answer the following question: *how can we accelerate Hamiltonian simulation while maintaining this algebraic and symplectic structure?* We address this by introducing a Lie-algebraic acceleration structure that turns the DAG-induced Hamiltonian generators into scan-compatible flow compositions.

Hamiltonian Drivers as Lie-group Members. For the initial physical state $z_0 = (q_0, p_0)$, we define Φ_m as the local Hamiltonian flow operator associated with the generators $\{g_m^{(k)}\}_{m=1}^{M}$, viewed as the elementary *Hamiltonian driver* of the trajectory. Specifically, the discrete time-evolution of Hamiltonian dynamics in (5) can be expressed by the successive composition of these operators:

$$z_0 \xrightarrow{\Phi_1} z_1 \xrightarrow{\Phi_2} z_2 \longrightarrow \cdots \longrightarrow z_{M-1} \xrightarrow{\Phi_M} z_M, \qquad (\Phi_M \circ \cdots \circ \Phi_1)$$

$$\begin{aligned} z_m &= \Phi_{g_m} z_{m-1}, \\ z_M &= \big(\Phi_M \circ \cdots \circ \Phi_1\big) z_0, \\ z_M &= (q_M, p_M). \end{aligned}$$

A central consequence of the introduced Neural-Poisson construction in Proposition 2.2 is that the Hamiltonian drivers $\{\Phi_m\}_{m=1}^{M}$ are not generic nonlinear transition maps. Rather, each driver can be viewed as an element of an algebraic structure called a *Lie group* Hilgert and Neeb [2012]. Indeed, for each $g_m \in \mathcal{G}^{(\leq K)}$, the Poisson action defines a finite dimensional Hamiltonian Lie operator:

$$\textbf{(Lie-algebraic Hamiltonian Driver):} \quad \mathcal{L}_{g_m,\leq r} f := \pi_{\leq r}\{f, g_m^{(k)}(\theta)\}, \quad \Phi_m := \exp\big(\epsilon \mathcal{L}_{g_m,\leq r}\big).$$

Thus, each Φ_m is a Lie-group element realized as a phase-space transformation, rather than a generic update function applied to z_{m-1}. Consequently, the trajectory-level map $\Phi_M \circ \cdots \circ \Phi_1$ is a series of group product in the Lie group with Hamiltonian flow transformations, where operator composition is recognized by the group operation. This Lie-group interpretation is made precise by the following proposition, which defines the DAG-indexed Lie-transform family obtained by exponentiating the truncated Poisson Lie operators associated with generators in $\mathcal{G}^{(\leq K)}$.

Proposition 3.1 (Neural–Poisson DAG Lie–Transform Family)**.** *For any element $g \in \mathcal{G}^{(\leq K)}$ and fixed $\epsilon > 0$, let $\mathcal{L}_{g,\leq r} : P_{\leq r} \to P_{\leq r}$ be the truncated Lie operator $\mathcal{L}_{g,\leq r} f := \pi_{\leq r}\big\{f, g\big\}$. Then, the family of DAG-indexed Hamiltonian drivers $(\mathscr{T}_{r,K}, \circ)$ is a **Lie group**:*

$$\mathscr{T}_{r,K} := \left\{ \Phi \;\middle|\; \Phi_g := \exp\big(\epsilon \mathcal{L}_{g,\leq r}\big) = \sum_{j=0}^{r} \frac{\epsilon^j}{j!} \mathcal{L}^j_{g,\leq r},\ \forall m \leq M \right\}. \tag{6}$$

The importance of this Lie-group formulation lies in the associativity of its intrinsic group operation. For any three Hamiltonian flow transformations $\Phi, \Psi, \Xi \in \mathscr{T}_{r,K}$, composition satisfies

$$\textbf{(Lie Group Associativity):} \qquad (\Phi \circ \Psi) \circ \Xi = \Phi \circ (\Psi \circ \Xi). \tag{7}$$

This property allows an ordered product of Hamiltonian drivers to be rebracketed without altering the resulting flow map. Consequently, the trajectory-level composition $\Phi_m \circ \cdots \circ \Phi_1$ can be evaluated as a scan-compatible prefix product on the Lie group, while preserving the temporal order of the drivers. This property plays a central role in the algebraic basis for parallel scan, where trajectory prefixes are computed as prefix products of Lie-group elements rather than by purely sequential modeling.

Neural–Poisson Lie Group Walk. The preceding construction characterizes each Hamiltonian driver Φ_{g_m} as a Lie-group element and identifies composition $\circ$ as the corresponding associative group operation. We now lift this *driver-level structure* to *trajectory-level evolution* by organizing a time ordered sequence of Hamiltonian drivers into a group-valued walk. This walk represents the accumulated evolution along the simulation horizon through recursive prefix products in $(\mathscr{T}_{r,K}, \circ)$.

Concretely, we define an *ordered Hamiltonian dictionary* $\mathbf{g} = [g_1, \ldots, g_m, \ldots, g_M]$ as a temporally ordered sequence of Hamiltonian generators, as illustrated in Fig. 1 *(left)*. In our implementation, each generator g_m is constructed as a sum of $|\mathcal{G}^{(\leq K)}|/M$ elements selected from the complete Neural–Poisson DAG $\mathcal{G}^{(\leq K)}$. The ordered application of these flow maps therefore defines a discrete walk on the Lie group of Hamiltonian flow transformations. This yields the following definition of the *Neural–Poisson Lie group walk*.

Definition 3.2 (Neural–Poisson Lie Group Walk)**.** Let us consider the Lie group $(\mathscr{T}_{r,K}, \circ)$, and $\mathbf{g}$ be an ordered dictionary of Hamiltonian generators. For $1 \leq m \leq M$, we define the *Neural–Poisson Lie Group walk* $\mathcal{W}_{\mathbf{g}} : \{0, 1, \ldots, M\} \to \mathscr{T}_{r,K}$ recursively by

$$\mathcal{W}_{\mathbf{g}}(0) := \mathbf{Id}, \quad \mathcal{W}_{\mathbf{g}}(m) := \Phi_m \circ \mathcal{W}_{\mathbf{g}}(m-1).$$

Intuitively, applying $\mathcal{W}_{\mathbf{g}}(m)$ to the initial phase-space state z_0 means successively flowing z_0 through the first m Hamiltonian drivers, $z_m = \mathcal{W}_{\mathbf{g}}(m)z_0 = \big(\Phi_{g_m} \circ \cdots \circ \Phi_{g_1}\big)z_0$. Thus, the physical state evolves by repeated Hamiltonian flows, while the composite transformation $\mathcal{W}_{\mathbf{g}}(m)$ itself traces a discrete walk on the Lie group $(\mathscr{T}_{r,K}, \circ)$. Fig. 1 *(right)* schematically illustrates the induced transition from z_0 to z_m. Following Definition 3.2, the generic Lie-group associativity in (7) can be rewritten directly in terms of the ordered members of the Neural-Poisson Lie group walk:

$$\big(\mathcal{W}_{\mathbf{g}}(m) \circ \mathcal{W}_{\mathbf{g}}(m-1)\big) \circ \mathcal{W}_{\mathbf{g}}(m-2) = \mathcal{W}_{\mathbf{g}}(m) \circ \big(\mathcal{W}_{\mathbf{g}}(m-1) \circ \mathcal{W}_{\mathbf{g}}(m-2)\big). \tag{8}$$

Remark. The use of (8) for prefix scan relies on representation-level closure, not on the formal associativity of arbitrary map composition. In SPS, both elementary drivers Φ_m and Lie product remain represented within the Neural-Poisson Lie structure, so the scan stays inside a structure-preserving family of Hamiltonian transformations. Existing unrestricted HNNs can compose flow maps abstractly, but without such a closed representation remain state-autoregressive and lack comparable algebraic guarantees for physical structure preservation.

Proposition 4.1 shows that the Lie-group walk $\mathcal{W}_{\mathbf{g}}(m)$ realizes an Lie-discretized Hamiltonian dynamics, where the discretization error is controlled by the prescribed truncation order $(i.e., r)$ while each prefix remains a composition of symplectic drivers. With the described algebraic property in (8), our next aim is to describe how to efficiently parallelize the computation of the trajectory via a symplectic version of *parallel prefix-scan* [Blelloch, 1990].

In parallelizing dynamics, prefix-scan algorithm computes all inclusive prefixes with respect to ordered Hamiltonian dictionary $\mathbf{g}$ within $O(\log M)$ parallel depth. Specifically, in our setting, these prefixes are precisely the Lie-group walk steps defined as follows:

$$\mathcal{W}_{\mathbf{g}}(m) := U_m \circ U_{m-1} \circ \cdots \circ U_1, \quad U_m = \Phi_m, \quad m = 1, \ldots, M. \tag{9}$$

By the associativity of the Lie-group operation in (8), all Lie-group walks can be evaluated simultaneously by organizing the prefix composites into a *balanced binary tree* as follows:

$$V_m^{(0)} := U_m, \quad V_m^{(1)} := U_{2m} \circ U_{2m-1}, \quad V_m^{(\ell)} := V_{2m}^{(\ell-1)} \circ V_{2m-1}^{(\ell-1)}, \quad \ell = 2, \ldots, \lceil \log_2 M \rceil.$$

where ℓ denotes the depth level of the balanced binary tree used in parallel scanning, $V_m^{(\ell)}$ represents the combined Lie flow of the contiguous block $U_{(2m-1)2^{\ell-1}+1}, \ldots, U_{2m2^{\ell-1}}$. By recursively evaluating these composites up to $\lceil \log_2 M \rceil$ yields every prefix in parallel.

Objective Function. We now return to the physical inverse problem introduced in (2), where the goal is to infer a Hamiltonian generator whose induced trajectory matches the observed evolution from a given initial condition:

$$\mathcal{L}(\theta) = \mathbb{E}_{\hat{z}_0, t}\left[\mathbb{E}_{\hat{z}_t}\left\|\mathcal{W}_{\mathbf{g}}(\lceil T/t \rceil)\hat{z}_0 - \hat{z}_t\right\|^2 \mid \hat{z}_0\right]. \tag{10}$$

As can be seen, the predicted state $z_t(\mathbf{g}) \coloneqq \mathcal{W}_{\mathbf{g}}(\lceil T/t \rceil)\hat{z}_0$ is obtained from the scan-evaluated Lie-group walk at the target horizon. Hence the loss remains a global trajectory-matching objective, while its trajectory map is realized through an $O(\log M)$-depth associative prefix computation rather than an $O(M)$-depth sequential trajectories required for existing HNN frameworks. In training, the same objective is applied to every sequentially re-indexed subtrajectory, including each consecutive transition. When all these terms vanish, the learned generators reproduce the target Hamiltonian vector field along the data and the walk preserves the target energy (Theorem A.2); otherwise the energy is preserved approximately, while the symplectic structure is preserved exactly.

4 Theoretical Findings

In this section, we rigorously establish two key theoretical results regarding the proposed *Neural–Poisson Lie–group walk*. First, we demonstrate its capacity for *universal approximation* of Hamiltonian functions (Proposition 4.1), providing explicit control over approximation errors. Second, we derive precise *generalization bounds* (Proposition 4.2), characterizing their dependence on the algebraic complexity inherent in the model.

Proposition 4.1. *Fix a finite horizon $T > 0$ and a compact set $K \subset \mathbb{R}^{2n}$. For exact Hamiltonians H lying in the bounded class $\mathcal{H}_{R,s} = \{H \in C^s : \|H\|_{C^s} \le R\}$ for $s > 2n$. Assume that our Neural–Poisson DAG can realize any polynomial of total degree $\le r$. Then,*

$$\sup_{H \in \mathcal{H}_{R,s}} \sup_{\hat{z}_0 \in K} \left\| \mathcal{W}_{\mathbf{g}}(M)(\hat{z}_0) - \hat{z}_{\lceil T/\epsilon \rceil} \right\| \lesssim \mathcal{O}\left(r^{1-s} + \epsilon^r / r!\right).$$

Proposition 4.1 shows that the Neural–Poisson DAG is an effective Hamiltonian approximator: increasing the polynomial degree r decreases the model approximation error r^{1-s}, while reducing the step size ϵ decreases the Lie–series truncation error $\epsilon^r/r!$. Thus both finer polynomial expressivity and temporal resolution jointly improve accuracy.

Proposition 4.2. *For any confidence level $\delta \in (0,1)$, with probability at least $1-\delta$ over an i.i.d. sample of size n, the worst-case difference between train and test loss satisfies*

$$\sup_{\mathcal{W}_{\mathbf{g}} \in \mathcal{F}_K} \left| \mathrm{Train}(\mathcal{W}_{\mathbf{g}}(M)) - \mathrm{Test}(\mathcal{W}_{\mathbf{g}}(M)) \right| \lesssim \mathcal{O}\left(\sqrt{\frac{M \log |\mathcal{G}^{(\le K)}|}{n}} + \sqrt{\frac{\log(1/\delta)}{n}} \right).$$

Complementing Proposition 4.1, this result quantifies how the same architectural choices enlarge the hypothesis class and hence affect generalization. In particular, increasing r and the DAG depth K increases the size of the generator family $\mathcal{G}^{(\le K)}$, which in turn amplifies the complexity term in the generalization bound. Together, Propositions 4.1 and 4.2 formalize a bias–variance trade-off in our design: enlarging the polynomial degree (and DAG) improves the ability to shadow the true Hamiltonian flow, but simultaneously increases $\log |\mathcal{G}^{(\le K)}|$ and therefore widens the potential generalization gap through the factor $\sqrt{M \log |\mathcal{G}^{(\le K)}|/n}$.

5 Related Work

Hamiltonian Neural Networks. Foundational structure-preserving approaches include Hamiltonian Neural Networks (HNNs) [Greydanus et al., 2019], which regress a scalar Hamiltonian whose canonical derivatives generate the dynamics, and SympNets [Jin et al., 2020], which directly parameterize symplectic maps through composition of symplectic layers. These classical models preserve important geometric structure, but their updates remain fundamentally sequential and operate through local residual minimization. Recent physics-informed approaches tailor training objectives to Hamiltonian systems by enforcing energy conservation or symplecticity directly in the loss or architecture. Beyond residual penalties, time-dependent symplectic neural flows parameterize Hamiltonian flow maps [Cañizares et al., 2024], and symplectic-loss training for HNNs improves long-horizon stability with a provable Hamiltonian [David and Méhats, 2023]. Discrete, variational-integrator flavored models on Lie groups preserve both group structure and symplectic form in learned maps [Duruisseaux et al., 2023]. Despite these advances, most methods still unroll dynamics sequentially and regularize *local* residuals or per-step updates, so their computational cost effectively grows linearly with trajectory length. In contrast, our goal is to exploit Hamiltonian structure for genuinely *fast*, parallelizable long-horizon simulation for $\mathcal{O}(\log M)$-depth complexity.

Neural Operator for Physical Dynamics. Operator-theoretic learning lifts nonlinear dynamics to linear evolution in an observable space. The Koopman Neural Operator (KNO) learns solution operators with mesh-free, long-horizon generalization. It has been applied to fast prediction of aerodynamic transonic buffet [Meng et al., 2024a]. Invertible-embedding approaches learn observable dictionaries with guaranteed reconstruction [Meng et al., 2024b]. Multiplicative-structure-preserving variants refine finite-dimensional approximations of observable algebras [Boullé and Colbrook, 2025]. Probabilistic and sequence-model extensions include KoVAE, which places a Koopman prior in a VAE for generative modeling of dynamical data [Naiman et al., 2024]. SKOLR approximates a structured Koopman operator via a linear-RNN factorization of lagged measurements [Zhang et al.,

Table 2: **Quantitative Comparison on Quantum Spin Dynamics (QSD).** Each column reports inference time, GPU memory, ADE/FDE, relative energy drift ΔE, and symplectic violation $\Delta\omega$ of 1000 time-steps. (A) Hamiltonian-based trajectory generators with explicit energy/symplectic preservation (B) operator-based generators without explicit preservation (C) black-box generators without preservation.

Integrator / Method	GPU Mem	Inference Time (s)	ADE / FDE ↓	ΔE ↓	$\Delta\omega$ ↓
(A) Hamiltonian Neural Network based Trajectory Generators with Energy / Symplectic Preservation					
HNN + Störmer-Verlet	1.2	3.47×10^{1}	$3.12\times10^{-3}/3.26\times10^{-3}$	$\mathbf{2.06\times10^{-3}}$	$\mathbf{5.31\times10^{-2}}$
HNN + Forest–Ruth 4th	1.5	4.01×10^{1}	$7.83\times10^{-4}/8.54\times10^{-4}$	$\mathbf{4.04\times10^{-4}}$	$\mathbf{2.73\times10^{-1}}$
HNN + Gauss–Legendre 4th	1.8	5.16×10^{1}	$5.76\times10^{-4}/7.12\times10^{-4}$	$\mathbf{1.15\times10^{-4}}$	$\mathbf{9.08\times10^{-1}}$
SymODEN [Zhong et al., 2020]	2.1	3.87×10^{1}	$8.64\times10^{-4}/9.93\times10^{-4}$	7.46×10^{-4}	2.63×10^{0}
SRNN [Chen et al., 2020]	2.5	4.53×10^{1}	$9.24\times10^{-4}/9.63\times10^{-4}$	8.13×10^{-4}	3.13×10^{0}
Nonsep-SNN [Xiong et al., 2021]	2.9	3.58×10^{1}	$7.56\times10^{-4}/8.63\times10^{-4}$	6.33×10^{-4}	2.23×10^{0}
SympNet [Jin et al., 2020]	2.3	3.21×10^{1}	$6.98\times10^{-4}/7.85\times10^{-4}$	5.74×10^{-4}	2.01×10^{0}
(B) Operator-based Trajectory Generators without Energy / Symplectic Preservation					
KoVAE [Naiman et al., 2024]	2.3	3.97×10^{-1}	$2.75\times10^{-4}/3.67\times10^{-4}$	2.31×10^{-3}	2.43×10^{1}
KoNODE [Bai and Ding, 2025]	2.7	5.04×10^{-1}	$3.22\times10^{-4}/4.67\times10^{-4}$	1.98×10^{-3}	1.76×10^{1}
(C) Black-box Trajectory Generators without Energy / Symplectic Preservation					
EqMotion [Xu et al., 2023]	24.6	6.88×10^{0}	$6.32\times10^{-5}/7.07\times10^{-5}$	$\mathbf{1.27\times10^{-1}}$	$\mathbf{5.23\times10^{1}}$
GeoTDM [Han et al., 2024]	18.5	1.83×10^{0}	$4.13\times10^{-5}/5.51\times10^{-5}$	$\mathbf{1.06\times10^{-1}}$	$\mathbf{4.88\times10^{1}}$
ET-SEED [Tie et al., 2025]	27.1	3.25×10^{0}	$3.96\times10^{-5}/5.62\times10^{-5}$	$\mathbf{9.97\times10^{-2}}$	$\mathbf{8.26\times10^{1}}$
Ours (SPS)	4.0	$\mathbf{3.31\times10^{-2}}$	$\mathbf{3.28\times10^{-5}/4.59\times10^{-5}}$	5.27×10^{-4}	4.02×10^{0}

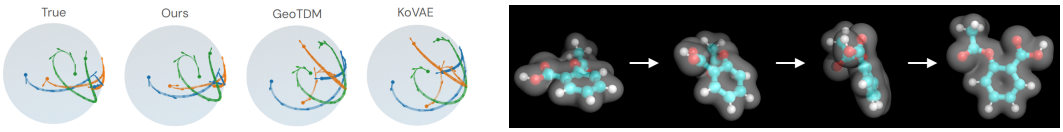


Figure 2: **Qualitative Results on Quantum Spin and Molecular Dynamics.** *(left)* Trajectory visualization of 4 spins on the Bloch sphere under the QSD Hamiltonian. The proposed method closely follows the ground-truth paths, whereas other models exhibit noticeable errors. *(right)* Snapshots of aspirin ($C_9H_8O_4$) along a 1000-step quantum simulation, illustrating rich internal motion while the overall molecular structure remains stable under the proposed integrator.

2025b]. KoNODE [Bai and Ding, 2025] introduces a Koopman-driven neural ODE where parameters evolve over time to capture nonstationary dynamics.

6 Experiments

We now present a series of experiments designed to evaluate the proposed framework across representative dynamical systems. Our goal is to test not only predictive accuracy but also the extent to which physical invariants and symplectic structures are preserved in long-horizon. Full details on numerical settings, benchmark specifications, dataset preparation, and additional statistics are deferred to Section A.6. Here we first highlight experiments on quantum spin dynamics, which serve as a canonical and physically rich testbed for assessing both accuracy and structure preservation.

Quantum Spin Dynamics (QSD). We consider a classical spin model widely used in magnetism and spintronics, where each spin is represented as a unit vector $\mathbf{S}_i \in \mathbb{CP}^1 \cong \mathbb{S}^2 \subset \mathbb{R}^3$. Spins in this system interact through a Hamiltonian incorporating multiple magnetic interactions, including Heisenberg exchange, Dzyaloshinskii–Moriya interactions (DMI), magnetocrystalline anisotropy, Zeeman coupling, long-range dipolar interactions, and next-nearest-neighbor exchange that accounts for frustration effects. We perform all experiments on systems consisting of 32 spins over trajectories spanning 1000 simulation steps. Table 2 compares our proposed method with three integrator categories, namely classical Hamiltonian-based integrators (A), neural operator-based approaches (B) and black-box methods (C). Methods in group (A) explicitly preserve physical invariants (highlighted in blue) but exhibit limited predictive accuracy, while methods in (B) achieve moderate accuracy with partial preservation of physical structure. The approaches in (C) may violate the physical invariants marked in red, despite their superior accuracy. In contrast, our method strikes an optimal balance by significantly enhancing predictive accuracy while robustly preserving essential physical invariants.

Molecular Dynamics: Small Molecules. We carry out molecular dynamics simulations for two prototypical molecular systems, methane (CH_4) and aspirin ($C_9H_8O_4$), with the OpenMM molecular dynamics simulation platform [Eastman et al., 2017]. We employed OpenFF or GAFF2 [Boothroyd et al., 2023, Wang et al., 2004] to parameterize interatomic interactions, and the classical Hamiltonian of molecular systems follows the standard AMBER functional form. All experiments were performed on systems with at most 40 atoms over trajectories spanning 1000 simulation steps. When traditional

Table 3: **Quantitative Comparison on Molecular Dynamics (MD).** We report inference runtime, per-molecule trajectory accuracy for methane, aspirin and peptide.

Method	Inference (s)	ADE / FDE (Methane) ↓	ADE / FDE (Aspirin) ↓	ADE / FDE (Peptide) ↓
(A) Black-box Trajectory Generators without Energy / Symplectic Preservation				
GeoTDM [Han et al., 2024]	1.91×10^{0}	$4.27\times10^{-5}/5.24\times10^{-5}$	$3.84\times10^{-5}/5.65\times10^{-5}$	$4.92\times10^{-5}/6.08\times10^{-5}$
ET-SEED [Tie et al., 2025]	3.09×10^{0}	$4.16\times10^{-5}/5.34\times10^{-5}$	$4.28\times10^{-5}/5.46\times10^{-5}$	$4.81\times10^{-5}/6.22\times10^{-5}$
(B) Machine Learning Force Field (MLFF) / Machine Interatomic Potential (MIP) Generators based Stochastic Integrator with Energy / Symplectic Preservation				
Störmer-Verlet + EquiformerV2 [Liao et al., 2024]	4.39×10^{1}	$7.54\times10^{-4}/8.23\times10^{-4}$	$8.39\times10^{-4}/9.36\times10^{-4}$	$9.12\times10^{-4}/1.05\times10^{-3}$
Störmer-Verlet + NequIP [Batzner et al., 2022]	5.87×10^{1}	$6.48\times10^{-4}/6.71\times10^{-4}$	$5.62\times10^{-4}/8.31\times10^{-4}$	$7.72\times10^{-4}/7.92\times10^{-4}$
Ours (SPS)	$\mathbf{4.39\times10^{-2}}$	$\mathbf{2.92\times10^{-5}/3.71\times10^{-5}}$	$\mathbf{3.36\times10^{-5}/4.27\times10^{-5}}$	$\mathbf{2.10\times10^{-5}/3.25\times10^{-5}}$

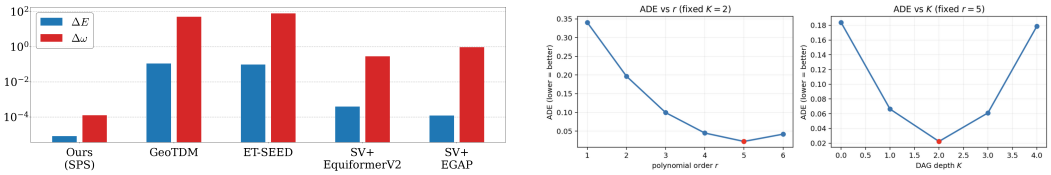


Figure 3: **Ablation on Physical Consistency and Hyperparameter Choices.** *(left)* Evaluation of physical consistency using the energy drift ΔE and the symplectic violation $\Delta\omega$, which measures distortion of the underlying Hamiltonian flow. The proposed method significantly reduces violations of physical invariants. *(right)* Analysis on the effect of the polynomial order r and the DAG depth K on the average displacement error (ADE), with $r = 5$ and $K = 2$ yield the best trade-off between accuracy and generalization.

symplectic integrators such as Störmer-Verlet are combined with existing MLFF generators in group (B) to simulate Hamiltonian dynamics, they require significantly large inference times, similar to many Hamiltonian neural network-based methods. In contrast, our SPS achieves remarkably faster inference while maintaining superior accuracy and preserving physical invariants.

Molecular Dynamics: Peptide. We conduct molecular dynamics simulations for the amyloid beta peptide fragment to assess a biomolecular setting with nontrivial conformational dynamics of peptides. We initialize the simulation from an experimentally resolved PDB (Protein Data Bank) structure [Kohno et al., 1996] and perform the simulation using the OpenMM molecular dynamics simulation platform under a classical Hamiltonian of molecular mechanics with an AMBER14 family force field [Maier et al., 2015]. All experiments were carried out on atomistic peptide systems comprising 155 atoms with trajectories of 1000 simulation steps. While conventional molecular dynamics simulations and learning-based force field models can incur substantial computational cost when applied to peptide-scale systems, particularly for long trajectories, our SPS enables significantly faster inference while preserving accurate modeling of peptide dynamics and key physical properties.

Ablation Study. We perform an ablation study with a comparative analysis of physical consistency across the models, evaluating energy conservation by using the energy drift ΔE and the symplectic violation $\Delta\omega$. The proposed method achieves a substantially lower energy drift and symplectic violation by factors of at least 10 and 10^3, respectively, compared to baseline models as shown in Figure 3 *(left)*. This substantial reduction suggests that the proposed method more faithfully respects the underlying Hamiltonian structure and leads to physically consistent long-term dynamics.

We further analyze the effect of model complexity on the approximation behavior of the proposed method. Motivated by the theoretical results in Section 4, we investigate the trade-off between polynomial order r, DAG depth K, and approximation capability. As shown in Figure 3 *(right)*, increasing either the polynomial order r or the DAG depth K leads to a more complex DAG structure that improves expressivity while progressively degrading generalization performance beyond a moderate regime. Varying both r and K systematically reveals that the best balance arises at polynomial order $r = 5$ and DAG depth $K = 2$, achieving minimal prediction error while maintaining stable generalization performance, as indicated by the observed ablation trend in our experiments.

7 Conclusion

In this work, we introduced SPS, a Poisson-algebraic framework for accelerating neural Hamiltonian simulation while preserving symplectic structure. By organizing DAG-induced Hamiltonian generators into a Lie-group walk, SPS converts long-horizon evolution into an associative prefix product with $\mathcal{O}(\log M)$ parallel depth. Experiments on spin and molecular dynamics show improved predictive accuracy, faster inference, and reduced violations of conserved physical quantities, suggesting a scalable route toward structure-preserving scientific simulation.

Acknowledgments and Disclosure of Funding

This work was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP)-ICT Creative Consilience Program grant funded by the Korea government (MSIT) (IITP-2026-RS-2020-II201819). This work was also supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2026-25495269). This research was supported by the Basic Science Research Program through the NRF funded by the Korea government (MSIT) (RS-2026-25522655).

References

Ralph Abraham and Jerrold E. Marsden. *Foundations of Mechanics*. Benjamin/Cummings, Reading, MA, 2 edition, 1978. ISBN 978-0-8053-0102-2.

Amikam Aharoni. *Introduction to the Theory of Ferromagnetism*. Oxford University Press, 1st edition, 1996.

Vladimir I. Arnold. *Mathematical Methods of Classical Mechanics*, volume 60 of *Graduate Texts in Mathematics*. Springer, 2 edition, 1989. doi: 10.1007/978-1-4757-2063-1. URL https://doi.org/10.1007/978-1-4757-2063-1.

Hanru Bai and Weiyang Ding. KoNODE: Koopman-driven neural ordinary differential equations with evolving parameters for time series analysis. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pages 2484–2519. PMLR, 2025. URL https://proceedings.mlr.press/v267/bai25b.html.

Andreas Bartel, Michael Günther, Birgit Jacob, and Timo Reis. Operator splitting based dynamic iteration for linear differential-algebraic port-hamiltonian systems. *Numerische Mathematik*, 155 (1):1–34, 2023. doi: 10.1007/s00211-023-01369-5.

Simon Batzner, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P. Mailoa, Mordechai Kornbluth, Nicola Molinari, Tess E. Smidt, and Boris Kozinsky. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature Communications*, 13:2453, 2022. doi: 10.1038/s41467-022-29939-5. URL https://doi.org/10.1038/s41467-022-29939-5.

Guy E. Blelloch. Prefix sums and their applications. Technical Report CMU-CS-90-190, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, 1990. URL https://www.cs.cmu.edu/~guyb/papers/Ble93.pdf.

Simon Boothroyd, Pavan Kumar Behara, Owen C. Madin, David F. Hahn, Hyesu Jang, Vytautas Gapsys, et al. Development and benchmarking of Open Force Field 2.0.0: The Sage small molecule force field. *Journal of Chemical Theory and Computation*, 19(11):3251–3275, 2023. doi: 10.1021/acs.jctc.3c00039. URL https://doi.org/10.1021/acs.jctc.3c00039.

Nicolas Boullé and Matthew J. Colbrook. Multiplicative dynamic mode decomposition. *SIAM Journal on Applied Dynamical Systems*, 24(2):1945–1968, 2025. doi: 10.1137/24M166317X.

Priscilla Cañizares, Davide Murari, Carola-Bibiane Schönlieb, Ferdia Sherry, and Zakhar Shumaylov. Symplectic neural flows for modeling and discovery. *arXiv preprint arXiv:2412.16787*, 2024.

Zhengdao Chen, Jianyu Zhang, Martin Arjovsky, and Léon Bottou. Symplectic recurrent neural networks. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=BkgYPREtPr.

Ana Cannas da Silva. *Lectures on Symplectic Geometry*, volume 1764 of *Lecture Notes in Mathematics*. Springer, 2008. doi: 10.1007/978-3-540-45330-7. URL https://doi.org/10.1007/978-3-540-45330-7.

Tri Dao, Daniel Y. Fu, Stefano Ermon, Atri Rudra, and Christopher Ré. FlashAttention: Fast and memory-efficient exact attention with IO-awareness. In *Advances in Neural Information Processing Systems*, volume 35, 2022. URL https://openreview.net/forum?id=H4DqfPSibmx.

Marco David and Florian Méhats. Symplectic learning for hamiltonian neural networks. *Journal of Computational Physics*, 494:112495, 2023. doi: 10.1016/j.jcp.2023.112495. URL https://doi.org/10.1016/j.jcp.2023.112495.

Valentin Duruisseaux, Thai P. Duong, Melvin Leok, and Nikolay Atanasov. Lie group forced variational integrator networks for learning and control of robot systems. In *Proceedings of the 5th Learning for Dynamics and Control Conference*, volume 211 of *Proceedings of Machine Learning Research*, pages 731–744. PMLR, 2023. URL https://proceedings.mlr.press/v211/duruisseaux23a.html.

I. Dzyaloshinsky. A thermodynamic theory of "weak" ferromagnetism of antiferromagnetics. *Journal of Physics and Chemistry of Solids*, 4(4):241–255, 1958. doi: 10.1016/0022-3697(58)90076-3.

Peter Eastman, Jason Swails, John D. Chodera, Robert T. McGibbon, Yutong Zhao, Kyle A. Beauchamp, Lee-Ping Wang, Andrew C. Simmonett, Matthew P. Harrigan, Chaya D. Stern, Rafal P. Wiewiora, Bernard R. Brooks, and Vijay S. Pande. Openmm 7: Rapid development of high-performance algorithms for molecular dynamics. *PLoS Computational Biology*, 13(7): e1005659, 2017. doi: 10.1371/journal.pcbi.1005659.

Richard F. L. Evans, Weijia J. Fan, Phanwadee Chureemart, Thomas A. Ostler, Matthew O. A. Ellis, and Roy W. Chantrell. Atomistic spin model simulations of magnetic nanomaterials. *Journal of Physics: Condensed Matter*, 26(10):103202, 2014. doi: 10.1088/0953-8984/26/10/103202. URL https://doi.org/10.1088/0953-8984/26/10/103202.

Etienne Forest and Robert D. Ruth. Fourth-order symplectic integration. *Physica D: Nonlinear Phenomena*, 43(1):105–117, 1990. doi: 10.1016/0167-2789(90)90019-L.

Samuel Greydanus, Misko Dzamba, and Jason Yosinski. Hamiltonian neural networks. In *Advances in Neural Information Processing Systems*, volume 32, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/26cd8ecadce0d4efd6cc8a8725cbd1f8-Abstract.html.

Albert Gu, Karan Goel, and Christopher Ré. Efficiently modeling long sequences with structured state spaces. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=uYLFoz1vlAC.

Ernst Hairer, Christian Lubich, and Gerhard Wanner. *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*, volume 31 of *Springer Series in Computational Mathematics*. Springer, Berlin, Heidelberg, 2 edition, 2006. doi: 10.1007/3-540-30666-8. URL https://doi.org/10.1007/3-540-30666-8.

Jiaqi Han, Minkai Xu, Aaron Lou, Haotian Ye, and Stefano Ermon. Geometric trajectory diffusion models. In *Advances in Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=G5vpJ1kMSK.

Werner Heisenberg. Zur theorie des ferromagnetismus. *Zeitschrift für Physik*, 49:619–636, 1928. doi: 10.1007/BF01328601. URL https://doi.org/10.1007/BF01328601.

Joachim Hilgert and Karl-Hermann Neeb. *Structure and Geometry of Lie Groups*. Springer, 2012. doi: 10.1007/978-0-387-84794-8. URL https://doi.org/10.1007/978-0-387-84794-8.

W. Daniel Hillis and Guy L. Steele Jr. Data parallel algorithms. *Communications of the ACM*, 29 (12):1170–1183, December 1986. doi: 10.1145/7902.7903. URL https://doi.org/10.1145/7902.7903.

Pengzhan Jin, Zhen Zhang, Aiqing Zhu, Yifa Tang, and George Em Karniadakis. SympNets: Intrinsic structure-preserving symplectic networks for identifying hamiltonian systems. *Neural Networks*, 132:166–179, 2020. doi: 10.1016/j.neunet.2020.08.017. URL https://doi.org/10.1016/j.neunet.2020.08.017.

Toshiyuki Kohno, Kuniko Kobayashi, Tadakazu Maeda, Kazuki Sato, and Akihiko Takashima. Three-dimensional structures of the amyloid β peptide (25–35) in membrane-mimicking environment. *Biochemistry*, 35(50):16094–16104, 1996. doi: 10.1021/bi961598j. URL https://doi.org/10.1021/bi961598j.

Shuo-Hui Li, Chen-Xiao Dong, Linfeng Zhang, and Lei Wang. Neural canonical transformation with symplectic flows. *Physical Review X*, 10(2):021020, 2020. doi: 10.1103/PhysRevX.10.021020. URL https://doi.org/10.1103/PhysRevX.10.021020.

Yi-Lun Liao, Brandon Wood, Abhishek Das, and Tess Smidt. EquiformerV2: Improved equivariant transformer for scaling to higher-degree representations. In *International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=mCOBKZmrzD.

James A. Maier, Carmenza Martinez, Koushik Kasavajhala, Lauren Wickstrom, Kevin E. Hauser, and Carlos Simmerling. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *Journal of Chemical Theory and Computation*, 11(8):3696–3713, 2015. doi: 10.1021/acs.jctc.5b00255. URL https://doi.org/10.1021/acs.jctc.5b00255.

Jerrold E. Marsden and Tudor S. Ratiu. *Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems*, volume 17 of *Texts in Applied Mathematics*. Springer, New York, NY, 2 edition, 1999. doi: 10.1007/978-0-387-21792-5. URL https://doi.org/10.1007/978-0-387-21792-5.

Dusa McDuff and Dietmar Salamon. *Introduction to Symplectic Topology*. Oxford University Press, Oxford, 2 edition, 1998. ISBN 978-0-19-850451-7.

Deying Meng, Yiding Zhu, Jianchun Wang, and Yipeng Shi. Koopman neural operator approach to fast flow prediction of airfoil transonic buffet. *Physics of Fluids*, 36(7):075182, 2024a. doi: 10.1063/5.0217336.

Yuhuang Meng, Jianguo Huang, and Yue Qiu. Koopman operator learning using invertible neural networks. *Journal of Computational Physics*, 501:112795, 2024b. doi: 10.1016/j.jcp.2024.112795. URL https://doi.org/10.1016/j.jcp.2024.112795.

Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of machine learning*. MIT press, 2018.

T. Moriya. Anisotropic superexchange interaction and weak ferromagnetism. *Physical Review*, 120 (1):91–98, 1960. doi: 10.1103/PhysRev.120.91. URL https://doi.org/10.1103/PhysRev.120.91.

Ilan Naiman, N. Benjamin Erichson, Pu Ren, Michael W. Mahoney, and Omri Azencot. Generative modeling of regular and irregular time series data via Koopman VAEs. In *International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=eY7sLb0dVF.

Jimmy T. H. Smith, Andrew Warrington, and Scott W. Linderman. Simplified state space layers for sequence modeling. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=Ai8Hw3AXqks.

Chenrui Tie, Yue Chen, Ruihai Wu, Boxuan Dong, Zeyi Li, Chongkai Gao, and Hao Dong. ET-SEED: Efficient trajectory-level SE(3) equivariant diffusion policy. In *International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=OheAR2xrtb.

Loup Verlet. Computer experiments on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Physical Review*, 159(1):98–103, 1967. doi: 10.1103/PhysRev.159.98.

Junmei Wang, Romain M. Wolf, James W. Caldwell, Peter A. Kollman, and David A. Case. Development and testing of a general amber force field. *Journal of Computational Chemistry*, 25(9): 1157–1174, 2004. doi: 10.1002/jcc.20035.

Shiying Xiong, Yunjin Tong, Xingzhe He, Shuqi Yang, Cheng Yang, and Bo Zhu. Nonseparable symplectic neural networks. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=B5VvQrI49Pa.

Chenxin Xu, Robby T. Tan, Yuhong Tan, Siheng Chen, Yu Guang Wang, Xinchao Wang, and Yanfeng Wang. EqMotion: Equivariant multi-agent motion prediction with invariant interaction reasoning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1410–1420, 2023.

Changwei Zhang, Yang Zhong, Zhi-Guo Tao, Xinming Qin, Honghui Shang, Zhenggang Lan, Oleg V. Prezhdo, Xin-Gao Gong, Weibin Chu, and Hongjun Xiang. Advancing nonadiabatic molecular dynamics simulations in solids with E(3) equivariant deep neural hamiltonians. *Nature Communications*, 16:2033, 2025a. doi: 10.1038/s41467-025-57328-1. URL https://doi.org/10.1038/s41467-025-57328-1.

Yitian Zhang, Liheng Ma, Antonios Valkanas, Boris N. Oreshkin, and Mark Coates. SKOLR: Structured Koopman operator linear RNN for time-series forecasting. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pages 75734–75756. PMLR, 2025b. URL https://proceedings.mlr.press/v267/zhang25be.html.

Yaofeng Desmond Zhong, Biswadip Dey, and Amit Chakraborty. Symplectic ODE-net: Learning hamiltonian dynamics with control. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=ryxmb1rKDS.

A Appendix

The appendix supplements the main text with detailed theoretical and algorithmic developments that underpin our framework. We first establish the *Symplecticity* property, demonstrating that the proposed dynamics respect the canonical two–form both in the single–exponential setting and for the full Neural–Poisson Lie–group walk. This is followed by the *Neural Poisson Algebra and Composition DAG*, where we formalize the truncated polynomial Poisson algebra, construct the associated composition DAG, and characterize the resulting Lie–transform family. In *Generalization Bound*, we provide a statistical analysis based on geometric complexity measures, leading to bounds on the generalization gap. The subsequent *Universal Approximation* subsection proves that our parameterization can approximate any target Hamiltonian flow to arbitrary accuracy within the truncated polynomial class. Finally, *Algorithmic Details* describes the implementation of the parallel–scan composition procedure and its integration into the Neural–Poisson Lie–group walk.

Notation Table

Table 4: Summary of notation used throughout the paper.

Symbol	Meaning
(M, ω)	Symplectic manifold with symplectic 2-form ω
$\Pi = \omega^{-1}$	Poisson bivector associated with ω
$\{f, g\}$	Poisson bracket of $f, g \in C^\infty(M)$
$\kappa : U \subset M \to \mathbb{R}^{2n}$	Darboux chart
$z = (q, p)$	Canonical coordinates $(q_1, \dots, q_n, p_1, \dots, p_n)$
J	Canonical symplectic matrix $\begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}$
X_H	Hamiltonian vector field $\omega^{-1} dH = J \nabla H$
$F^t_{X_H}$	Time-t Hamiltonian flow of X_H
Poly	Polynomial ring in (q, p)
Poly^d	Homogeneous polynomials of degree d
$F_r\mathrm{Poly}$	Space of polynomials of total degree $\leq r$
$\pi_{\leq r}$	Projection onto degree-$\leq r$ polynomial subspace
m_θ	Projected multiplication $m_\theta(f, g) = \pi_{\leq r}(fg)$
$\{\cdot, \cdot\}_\theta$	Projected Poisson bracket $\pi_{\leq r}\{f, g\}$
L_H	Projected Lie derivative $f \mapsto \{f, H\}_\theta$
$P_{\leq r}$	Space of degree-$\leq r$ polynomials
$\mathcal{G}^{(\leq K)}$	Depth-$\leq K$ Neural–Poisson Composition DAG
$L_{g,\leq r}$	Truncated Lie operator $\pi_{\leq r}\{f, g\}$
$\mathfrak{t}_{r,K}$	Lie algebra spanned by $\{L_{g,\leq r} : g \in \mathcal{G}^{(\leq K)}\}$
$\mathscr{T}_{r,K}$	Lie group of truncated flows $\exp(\epsilon L_{g,\leq r})$
$\Phi_{\epsilon,g}$	ϵ-time flow generated by g
$W_g(m)$	Neural–Poisson Lie–group walk after m steps
U_m	m-th elementary Lie group update Φ_{g_m}
$V_m^{(\ell)}$	Balanced binary-tree scan composition node
C^s, H^s	Function spaces of order s (smooth / Sobolev)
$\|\cdot\|_{C^s}$	C^s norm
μ_ω	Liouville measure $\omega^{\wedge n}/n!$

A.1 Symplecticity

In this subsection, we verify that the proposed framework preserves the fundamental symplectic structure of Hamiltonian mechanics, starting from the single–exponential case and extending to the full Neural–Poisson Lie–group walk.

Proposition A.1. *Let (M, ω) be a symplectic manifold with Poisson bracket $\{\cdot,\cdot\}$. Fix $H \in C^\infty(M)$ and write $X_H := \omega^{-1} dH$ for its Hamiltonian vector field with flow $F^t_{X_H}$. Assume the Lie–scan composition factorizes as a single exponential*

$$\Psi_{\Delta t} = \exp\big(\Delta t L_H\big), \quad L_H f := \{f, H\},$$

for some $\Delta t \in I$. Then $\Psi_{\Delta t} = F^{\Delta t}_{X_H}$, hence the discrete update $z_{k+1} = \Psi_{\Delta t}(z_k)$ coincides with sampling the continuous solution of $\dot z_t = X_H(z_t)$ at times $t = k\Delta t$.

Proof. Fix $\Delta t \in I$ and work in a Darboux chart, where $X_H = J\nabla H$ and $\{f, g\} = \nabla f^\top J \nabla g$. As in the Lie-series convention, an operator acting on functions defines a point map through its action on the coordinate functions $z_1, \dots, z_{2n}$, and the convergence of the exponential series is part of the hypothesis:

$$\Psi_{\Delta t}(z) := \Big(\big(\exp(\Delta t L_H) z_1\big)(z), \dots, \big(\exp(\Delta t L_H) z_{2n}\big)(z)\Big). \tag{11}$$

We compare two operator families on $C^\infty(M)$, the Koopman pullback and the Lie series,

$$U_t f := f \circ F^t_{X_H}, \qquad V_t f := \exp(t L_H) f = \sum_{k\ge 0} \frac{t^k}{k!} L_H^k f. \tag{12}$$

For every $f \in C^\infty(M)$, $t \in I$ and $z \in M$, the chain rule and the flow equation give

$$\begin{aligned}
\frac{d}{dt}\big(U_t f\big)(z) &= \nabla f\big(F^t_{X_H}(z)\big)^\top \frac{d}{dt} F^t_{X_H}(z) = \nabla f\big(F^t_{X_H}(z)\big)^\top J \nabla H\big(F^t_{X_H}(z)\big) \\
&= \sum_{i=1}^{n} \Big(\partial_{q_i} f\, \partial_{p_i} H - \partial_{p_i} f\, \partial_{q_i} H\Big)\big(F^t_{X_H}(z)\big) \\
&= \{f, H\}\big(F^t_{X_H}(z)\big) = \big(L_H f\big)\big(F^t_{X_H}(z)\big) = \big(U_t L_H f\big)(z),
\end{aligned} \tag{13}$$

and differentiating the series in (12) term by term gives

$$\frac{d}{dt} V_t f = \sum_{k\ge 1} \frac{k\, t^{k-1}}{k!} L_H^k f = \sum_{k\ge 0} \frac{t^k}{k!} L_H^k \big(L_H f\big) = V_t L_H f = L_H V_t f, \tag{14}$$

where the last equality holds because V_t is a power series in L_H. Fix $t \in I$. Applying (13) to the function $V_s f$ and using (14),

$$\begin{aligned}
\frac{d}{ds}\Big(U_{t-s} V_s f\Big) &= -\,U_{t-s} L_H\big(V_s f\big) + U_{t-s}\Big(\frac{d}{ds} V_s f\Big) \\
&= -\,U_{t-s} L_H V_s f + U_{t-s} L_H V_s f = 0, \qquad s \in [0, t].
\end{aligned} \tag{15}$$

The function $s \mapsto U_{t-s} V_s f$ is therefore constant on $[0, t]$, and evaluating it at both end points with $U_0 = V_0 = \mathrm{Id}$ yields

$$U_t f = U_{t-0}\, V_0 f = U_{t-t}\, V_t f = V_t f. \tag{16}$$

Taking $t = \Delta t$ and $f = z_i$ in (16) and comparing with (11), for $i = 1, \dots, 2n$,

$$\big(F^{\Delta t}_{X_H}(z)\big)_i = \big(z_i \circ F^{\Delta t}_{X_H}\big)(z) = \big(U_{\Delta t} z_i\big)(z) = \big(V_{\Delta t} z_i\big)(z) = \big(\Psi_{\Delta t}(z)\big)_i, \tag{17}$$

hence $\Psi_{\Delta t} = F^{\Delta t}_{X_H}$. By (17), induction on k, and the group property $F^t_{X_H} \circ F^s_{X_H} = F^{t+s}_{X_H}$ of autonomous flows,

$$z_{k+1} = \Psi_{\Delta t}(z_k) = F^{\Delta t}_{X_H}(z_k) = F^{\Delta t}_{X_H}\big(F^{k\Delta t}_{X_H}(z_0)\big) = F^{(k+1)\Delta t}_{X_H}(z_0),$$

which is the solution of $\dot z_t = X_H(z_t)$ sampled at $t = (k+1)\Delta t$. □

Theorem A.2 (Neural–Poisson Lie–group walk is symplectic)**.** *For any ordered dictionary* $\mathbf{g}$*, let* $\mathscr{T}_{r,K}$ *be the inclusive prefixes of the Neural–Poisson Lie–group walk, which preserves the symplectic structure and the phase-space volume:*

$$\underbrace{\mathcal{W}_{\mathbf{g}}^{*}\omega = \omega}_{\textit{structure}}, \qquad \underbrace{\det D\mathcal{W}_{\mathbf{g}} = 1}_{\textit{volume}}, \quad \forall m \le M, \quad z \in \mathcal{M}.$$

In the ideal zero-loss regime of trajectory matching, where the learned generators reproduce the target vector field on the region visited by the walk, the walk additionally preserves the target energy:

$$\underbrace{H\big(\mathcal{W}_{\mathbf{g}} z\big) = H(z)}_{\textit{energy}}, \quad \forall m \le M.$$

Consequently the discrete trajectory $z_m = \mathcal{W}_{\mathbf{g}}(m) z_0$ *preserves the symplectic form and the Liouville measure* $\mu_\omega = \omega^{\wedge n}/n!$ *at every step, and the energy at every step in the ideal zero-loss regime.*

Proof. The argument extends Proposition A.1 from a *single* exponential factor to the *ordered product* that defines the Neural–Poisson Lie–group walk. Let the ordered dictionary be $\mathbf{g} = (g_1, \dots, g_M) \subset \mathrm{Poly}_\theta^{(r)}$ with step sizes $\Delta_1, \dots, \Delta_M > 0$, and set $\Phi_m := \exp(\Delta_m L_{g_m})$. Proposition A.1 identifies each driver with an exact Hamiltonian flow,

$$\Phi_m = F_{X_{g_m}}^{\Delta_m}, \qquad X_{g_m} = J\nabla g_m, \qquad DX_{g_m} = J\,\nabla^2 g_m, \tag{18}$$

whose Jacobian solves the variational equation

$$\frac{d}{dt} DF_{X_{g_m}}^{t}(z) = DX_{g_m}\big(F_{X_{g_m}}^{t}(z)\big)\, DF_{X_{g_m}}^{t}(z), \qquad DF_{X_{g_m}}^{0} = I_{2n}. \tag{19}$$

In Darboux coordinates $\omega_z(u,v) = u^\top J v$, hence $(\Phi^*\omega)_z(u,v) = u^\top D\Phi(z)^\top J\, D\Phi(z)\, v$ and

$$\Phi^*\omega = \omega \quad \Longleftrightarrow \quad D\Phi(z)^\top J\, D\Phi(z) = J \quad \text{for all } z. \tag{20}$$

Suppressing the arguments z and $F_{X_{g_m}}^{t}(z)$, the variational equation (19), $DX_{g_m} = J\nabla^2 g_m$ from (18), the symmetry of $\nabla^2 g_m$, and $J^\top J = -JJ = I_{2n}$ give

$$\begin{aligned}
\frac{d}{dt}\Big(DF_{X_{g_m}}^{t\,\top}\, J\, DF_{X_{g_m}}^{t}\Big) &= \Big(\frac{d}{dt} DF_{X_{g_m}}^{t}\Big)^\top J\, DF_{X_{g_m}}^{t} + DF_{X_{g_m}}^{t\,\top} J\, \frac{d}{dt} DF_{X_{g_m}}^{t} \\
&= DF_{X_{g_m}}^{t\,\top}\Big(DX_{g_m}^\top J + J\, DX_{g_m}\Big) DF_{X_{g_m}}^{t} \\
&= DF_{X_{g_m}}^{t\,\top}\Big(\nabla^2 g_m\, J^\top J + JJ\,\nabla^2 g_m\Big) DF_{X_{g_m}}^{t} \\
&= DF_{X_{g_m}}^{t\,\top}\Big(\nabla^2 g_m - \nabla^2 g_m\Big) DF_{X_{g_m}}^{t} = 0.
\end{aligned} \tag{21}$$

Integrating (21) from 0 to Δ_m with the initial condition in (19), and using (20),

$$D\Phi_m^\top J\, D\Phi_m = DF_{X_{g_m}}^{\Delta_m\,\top} J\, DF_{X_{g_m}}^{\Delta_m} = DF_{X_{g_m}}^{0\,\top} J\, DF_{X_{g_m}}^{0} = J, \qquad \Phi_m^*\omega = \omega. \tag{22}$$

Jacobi's formula for the derivative of a determinant, the variational equation (19), and $X_{g_m} = (\nabla_p g_m, -\nabla_q g_m)$ give

$$\begin{aligned}
\frac{d}{dt} \det DF_{X_{g_m}}^{t}(z) &= \det DF_{X_{g_m}}^{t}(z)\ \operatorname{tr}\Big(\big(DF_{X_{g_m}}^{t}(z)\big)^{-1} \frac{d}{dt} DF_{X_{g_m}}^{t}(z)\Big) \\
&= \det DF_{X_{g_m}}^{t}(z)\ \operatorname{tr} DX_{g_m}\big(F_{X_{g_m}}^{t}(z)\big) \\
&= \det DF_{X_{g_m}}^{t}(z)\ \sum_{i=1}^{n}\Big(\partial_{q_i}\partial_{p_i} g_m - \partial_{p_i}\partial_{q_i} g_m\Big)\big(F_{X_{g_m}}^{t}(z)\big) = 0,
\end{aligned}$$

hence

$$\det D\Phi_m(z) = \det DF_{X_{g_m}}^{\Delta_m}(z) = \det DF_{X_{g_m}}^{0}(z) = 1. \tag{23}$$

Each driver conserves its own generator, since $\{g_m, g_m\} = -\{g_m, g_m\} = 0$ by antisymmetry. For the target Hamiltonian H, we work in the ideal zero-loss regime of trajectory matching: every consecutive transition of the sequentially re-indexed objective is reproduced exactly, and the learned generators reproduce the target vector field on the region visited by the walk. The antisymmetry of J, $u^\top J u = 0$, then gives

$$J\nabla g_m(z) = J\nabla H(z) \quad \Longrightarrow \quad \{H, g_m\}(z) = \nabla H(z)^\top J \nabla g_m(z) = \nabla H(z)^\top J \nabla H(z) = 0. \tag{24}$$

For $f = g_m$, and for $f = H$ under (24),

$$\frac{d}{dt} f\big(F^t_{X_{g_m}}(z)\big) = \nabla f\big(F^t_{X_{g_m}}(z)\big)^\top J \nabla g_m\big(F^t_{X_{g_m}}(z)\big) = \{f, g_m\}\big(F^t_{X_{g_m}}(z)\big) = 0, \tag{25}$$

and integrating (25) from 0 to Δ_m,

$$H(\Phi_m z) = H(z), \qquad g_m(\Phi_m z) = g_m(z). \tag{26}$$

The inclusive prefixes satisfy $\mathcal{W}_{\mathbf{g}}(0) = \mathbf{Id}$ and $\mathcal{W}_{\mathbf{g}}(m) = \Phi_m \circ \mathcal{W}_{\mathbf{g}}(m-1)$, and the chain rule gives

$$D\mathcal{W}_{\mathbf{g}}(m)(z) = D\Phi_m\big(\mathcal{W}_{\mathbf{g}}(m-1)z\big)\, D\mathcal{W}_{\mathbf{g}}(m-1)(z). \tag{27}$$

Writing $D\Phi_m$ for $D\Phi_m\big(\mathcal{W}_{\mathbf{g}}(m-1)z\big)$ and $D\mathcal{W}_{\mathbf{g}}(m)$ for $D\mathcal{W}_{\mathbf{g}}(m)(z)$, the identities (27), (22), (23) and (26) yield, for every $1 \le m \le M$,

$$\begin{aligned} D\mathcal{W}_{\mathbf{g}}(m)^\top J\, D\mathcal{W}_{\mathbf{g}}(m) &= D\mathcal{W}_{\mathbf{g}}(m-1)^\top \big(D\Phi_m^\top J\, D\Phi_m\big) D\mathcal{W}_{\mathbf{g}}(m-1) \\ &= D\mathcal{W}_{\mathbf{g}}(m-1)^\top J\, D\mathcal{W}_{\mathbf{g}}(m-1) \\ &= \cdots = D\mathcal{W}_{\mathbf{g}}(0)^\top J\, D\mathcal{W}_{\mathbf{g}}(0) = J, \end{aligned} \tag{28}$$

$$\begin{aligned} \det D\mathcal{W}_{\mathbf{g}}(m) &= \det D\Phi_m \ \det D\mathcal{W}_{\mathbf{g}}(m-1) = \det D\mathcal{W}_{\mathbf{g}}(m-1) \\ &= \cdots = \det D\mathcal{W}_{\mathbf{g}}(0) = 1, \end{aligned} \tag{29}$$

$$\begin{aligned} H\big(\mathcal{W}_{\mathbf{g}}(m)z\big) &= H\big(\Phi_m\big(\mathcal{W}_{\mathbf{g}}(m-1)z\big)\big) = H\big(\mathcal{W}_{\mathbf{g}}(m-1)z\big) \\ &= \cdots = H\big(\mathcal{W}_{\mathbf{g}}(0)z\big) = H(z). \end{aligned} \tag{30}$$

By (20), (28) is the coordinate form of $\mathcal{W}_{\mathbf{g}}(m)^*\omega = \omega$; equivalently, since pull-backs reverse the order of composition,

$$\begin{aligned} \mathcal{W}_{\mathbf{g}}(m)^*\omega &= \Phi_1^*\big(\Phi_2^* \cdots \big(\Phi_m^*\omega\big)\big) = \Phi_1^*\big(\Phi_2^* \cdots \big(\Phi_{m-1}^*\omega\big)\big) = \cdots = \omega, \\ \mathcal{W}_{\mathbf{g}}(m)^*\mu_\omega &= \frac{1}{n!}\big(\mathcal{W}_{\mathbf{g}}(m)^*\omega\big)^{\wedge n} = \frac{1}{n!}\,\omega^{\wedge n} = \mu_\omega. \end{aligned} \tag{31}$$

For the discrete trajectory $z_m := \mathcal{W}_{\mathbf{g}}(m)z_0$ with $z_0 \in \mathcal{M}$, the identities (28), (29) and (31) establish the preservation of the symplectic form and the Liouville measure $\mu_\omega = \omega^{\wedge n}/n!$ for every ordered dictionary, and (30) establishes $H(z_m) = H(z_0)$ at every step $1 \le m \le M$ in the ideal zero-loss regime. Outside this regime, the trajectory-matching loss and the polynomial truncation error are nonzero, and (24) holds only approximately. Since $\nabla H^\top J \nabla H = 0$, the energy changes along each driver at the rate (32), and integrating over each driver and summing along the walk bound the drift by the vector-field mismatch:

$$\frac{d}{dt} H\big(F^t_{X_{g_j}}(z)\big) = \nabla H^\top J \nabla g_j\big(F^t_{X_{g_j}}(z)\big) = \nabla H^\top J\big(\nabla g_j - \nabla H\big)\big(F^t_{X_{g_j}}(z)\big), \tag{32}$$

$$\begin{aligned} \big|H\big(\mathcal{W}_{\mathbf{g}}(m)z\big) - H(z)\big| &\le \sum_{j=1}^{m} \Big|H\big(\Phi_j\big(\mathcal{W}_{\mathbf{g}}(j-1)z\big)\big) - H\big(\mathcal{W}_{\mathbf{g}}(j-1)z\big)\Big| \\ &\le \sum_{j=1}^{m} \Delta_j\, \|\nabla H\|_{L^\infty}\, \|J\nabla g_j - J\nabla H\|_{L^\infty}. \end{aligned} \tag{33}$$

The bound (33) vanishes in the ideal zero-loss regime, whereas (28) and (29) hold exactly for every ordered dictionary. □

A.2 Neural Poisson Algebra and Composition DAG

Before presenting the formal construction, we briefly outline the algebraic setting that underlies the proposed framework, providing the basic structures and notation used throughout the paper.

Let $(\mathcal{M}, \omega)$ be a symplectic manifold with Poisson bivector $\Pi = \omega^{-1}$ with $\{f, g\} = \Pi(df, dg)$. Given a Darboux chart $\kappa : U \subset M \to \mathbb{R}^{2n}$ for explicit coordinates, though everything below is written intrinsically in terms of Π and pull-backs. Let us set polynomial space and their d-order subspace:

$$\mathrm{Poly} := \mathbb{R}[q_1, \ldots, q_n, p_1, \ldots, p_n], \qquad \mathrm{Poly}_d := \{f \in \mathrm{Poly} \mid \deg f = d\}.$$

This structure naturally admits the filtration $F^r\mathrm{Poly} := \bigoplus_{d \le r} \mathrm{Poly}_d$ for any $r \in \mathbb{N}$. While $\deg(fg) = \deg f + \deg g$ and $\deg\{f, g\} \le \deg f + \deg g - 2$, both the product and the Poisson bracket preserve the filtration. For the fixed truncation order $r \ge 1$ with $\pi_{\le r} : \mathrm{Poly} \to F^r\mathrm{Poly}$ of the degree-$\le r$ projection, note that $\pi^2_{\le r} = \pi_{\le r}$. Now, we introduce the projected bilinear maps

$$m_\theta(f, g) := \pi_{\le r}(fg), \qquad \{f, g\}_\theta := \pi_{\le r}\big(\{f, g\}\big), \qquad f, g \in F^r\mathrm{Poly}.$$

Consequently, the following triplet $\mathrm{Poly}^{(r)}_\theta := \big(F^r\mathrm{Poly}, m_\theta, \{\cdot, \cdot\}_\theta\big)$ is called a *projected Poisson algebra*. The product m_θ is commutative, associative, and unital (with unit 1) within $F^r\mathrm{Poly}$. The projected Poisson bracket $\{\cdot, \cdot\}_\theta$ is bilinear and antisymmetric, and its Jacobiator equals $\pi_{\le r}(\mathsf{J}(f, g, h))$, so it vanishes whenever the unprojected Jacobiator exceeds degree r. Moreover, the projected Leibniz rule holds: $\{f, m_\theta(g, h)\}_\theta = m_\theta(\{f, g\}_\theta, h) + m_\theta(g, \{f, h\}_\theta)$.s Note that every $H \in F^r\mathrm{Poly}$ induces a Hamiltonian derivation $L_H : f \longmapsto \{f, H\}_\theta$ and $H \mapsto L_H$ is a Lie representation: $[L_{H_1}, L_{H_2}] = L_{\{H_1, H_2\}_\theta}$. This shows that the following space forms a finite-dimensional Lie algebra.

$$\mathfrak{g}^{(r)}_\theta := \{L_H \mid H \in F^r\mathrm{Poly}\}$$

Exponentiating L_H yields a canonical transformation $\exp(\Delta t L_H)$ that coincides with the time-Δt flow of the projected Hamiltonian vector field whenever the Lie–scan collapses to a single exponential.

We now formalize this observation by characterizing the full family of such truncated Lie–generated flows within our DAG framework.

Proposition A.3 (Neural–Poisson DAG Lie–Transform Family)**.** *Fix integers $r \ge 1$ and $K \ge 0$ and recall the Neural–Poisson Composition Graph $\mathcal{G}^{(\le K)}$ in Definition 2.2. For any element $g \in \mathcal{G}^{(\le K)}$ let $\mathcal{L}_{g,\le r} : P_{\le r} \to P_{\le r}$ denote the truncated Lie operator $\mathcal{L}_{g,\le r} f := \pi_{\le r}\{f, g\}$. Then the family of DAG-indexed Lie transforms $\mathscr{T}_{r,K}$ is a* ***Lie group***[1]*:*

$$\mathscr{T}_{r,K} := \Big\{\Phi_{\epsilon,g} := \exp\big(\epsilon \mathcal{L}_{g,\le r}\big) = \sum_{j=0}^{r} \frac{\epsilon^j}{j!} \mathcal{L}^j_{g,\le r} \;\Big|\; g \in \mathcal{G}^{(\le K)}\Big\}. \tag{34}$$

Proof of Proposition 3.1. Throughout, we adopt the symbols introduced just above: $P_{\le r}$, the projected Poisson bracket $\{\cdot, \cdot\}_\theta$, and the truncated Lie operators $\mathcal{L}_{g,\le r}$. The monomials $q^{\boldsymbol{\alpha}} p^{\boldsymbol{\beta}}$ with $|\boldsymbol{\alpha}| + |\boldsymbol{\beta}| \le r$ form a basis of $P_{\le r}$, hence $N := \dim P_{\le r} = \binom{2n+r}{r}$. For $f, h \in P_{\le r}$, $g, g' \in \mathcal{G}^{(\le K)}$ and $a, b \in \mathbb{R}$, bilinearity of the bracket and linearity of $\pi_{\le r}$ give

$$\mathcal{L}_{g,\le r}(af + bh) = \pi_{\le r}\{af + bh,\, g\} = \pi_{\le r}\big(a\{f, g\} + b\{h, g\}\big) = a\,\mathcal{L}_{g,\le r} f + b\,\mathcal{L}_{g,\le r} h, \tag{35}$$

$$\mathcal{L}_{ag+bg',\le r} f = \pi_{\le r}\{f,\, ag + bg'\} = \pi_{\le r}\big(a\{f, g\} + b\{f, g'\}\big) = a\,\mathcal{L}_{g,\le r} f + b\,\mathcal{L}_{g',\le r} f. \tag{36}$$

By (35), $\mathcal{L}_{g,\le r} \in \mathrm{End}(P_{\le r}) \cong \mathfrak{gl}_N(\mathbb{R})$ after fixing the monomial basis; by (36), $g \mapsto \mathcal{L}_{g,\le r}$ is linear. For the untruncated operators $\mathcal{L}_g f := \{f, g\}$, antisymmetry and the Jacobi identity give, for every f,

$$\begin{aligned}[\mathcal{L}_{g_1}, \mathcal{L}_{g_2}]f &= \{\{f, g_2\}, g_1\} - \{\{f, g_1\}, g_2\} = \{g_1, \{g_2, f\}\} + \{g_2, \{f, g_1\}\} \\ &= -\{f, \{g_1, g_2\}\} = \{f, \{g_2, g_1\}\} = \mathcal{L}_{\{g_2, g_1\}} f,\end{aligned} \tag{37}$$

hence commutators of Lie operators are again Lie operators, and the generators of the Neural–Poisson DAG are closed under the projected bracket by construction. After truncation, the commutators of the

operators $\mathcal{L}_{g,\le r}$ remain linear maps on $P_{\le r}$, and the Lie algebra spanned by the truncated operators,

$$\mathfrak{t}_{r,K} := \operatorname{span}\Big\{ \big[\mathcal{L}_{g_1,\le r}, \big[\cdots, \big[\mathcal{L}_{g_{j-1},\le r}, \mathcal{L}_{g_j,\le r}\big]\cdots\big]\big] \;\Big|\; j \ge 1,\ g_i \in \mathcal{G}^{(\le K)} \Big\} \subset \mathfrak{gl}_N(\mathbb{R}), \quad (38)$$

is closed under the commutator and satisfies $\dim \mathfrak{t}_{r,K} \le N^2 < \infty$. By the correspondence between Lie subalgebras and connected Lie subgroups [Hilgert and Neeb, 2012, Ch. II], there exists a unique connected Lie subgroup $\mathscr{T}_{r,K} \subset \mathrm{GL}_N(\mathbb{R})$ whose Lie algebra is (38), namely the subgroup generated by $\exp(\mathfrak{t}_{r,K})$:

$$\mathscr{T}_{r,K} = \Big\{ \exp(X_j)\circ\cdots\circ\exp(X_1) \;\Big|\; j \in \mathbb{N},\ X_1,\dots,X_j \in \mathfrak{t}_{r,K} \Big\}, \qquad \dim \mathscr{T}_{r,K} = \dim \mathfrak{t}_{r,K}. \quad (39)$$

The group operations are the restrictions of matrix multiplication, which is the composition $\circ$ of linear maps on $P_{\le r}$, and of matrix inversion; both are smooth.

Fix $g \in \mathcal{G}^{(\le K)}$ and $\epsilon \in \mathbb{R}$. Since $\epsilon\mathcal{L}_{g,\le r} \in \mathfrak{t}_{r,K}$, (39) gives $\Phi_{\epsilon,g} = \exp(\epsilon\mathcal{L}_{g,\le r}) \in \mathscr{T}_{r,K}$. Whenever $\mathcal{L}_{g,\le r}$ is nilpotent of order $r+1$, the exponential series terminates:

$$\exp(\epsilon\mathcal{L}_{g,\le r}) = \sum_{j=0}^{r} \frac{\epsilon^j}{j!}\mathcal{L}_{g,\le r}^{j} + \sum_{j\ge r+1} \frac{\epsilon^j}{j!}\mathcal{L}_{g,\le r}^{j-r-1}\,\mathcal{L}_{g,\le r}^{r+1} = \sum_{j=0}^{r} \frac{\epsilon^j}{j!}\mathcal{L}_{g,\le r}^{j}. \quad (40)$$

This is the case for every homogeneous generator g of degree $d \ne 2$. For homogeneous f of degree k, $\{f,g\}$ is homogeneous of degree $k+d-2$, constants are annihilated, and $\pi_{\le r}$ removes every component of degree larger than r, hence

$$\begin{aligned} d = 1: &\qquad \mathcal{L}_{g,\le r}\big(\mathrm{Poly}_k\big) \subset \mathrm{Poly}_{k-1} &&\Longrightarrow \quad \mathcal{L}_{g,\le r}^{r+1} = 0, \\ d \ge 3: &\qquad \mathcal{L}_{g,\le r}^{j}\big(\mathrm{Poly}_k\big) \subset \mathrm{Poly}_{k+j(d-2)},\ \ k \ge 1 &&\Longrightarrow \quad \mathcal{L}_{g,\le r}^{r} = 0, \end{aligned} \quad (41)$$

since $k + j(d-2) > r$ for $j = r$ and $k \ge 1$; in both cases (40) applies. Conversely, the subgroup generated by the one-parameter subgroups $\{\exp(s\mathcal{L}_{g,\le r}) : s \in \mathbb{R}\}$, $g \in \mathcal{G}^{(\le K)}$, is a connected Lie subgroup of $\mathscr{T}_{r,K}$ whose Lie algebra contains every $\mathcal{L}_{g,\le r}$ together with all their iterated commutators [Hilgert and Neeb, 2012]; its Lie algebra is therefore (38), and the two connected subgroups coincide. The neural coefficients are real-valued, and rescaling a DAG node rescales one of its leaves, $a\,\pi_{\le r}(f \star_\theta h) = \pi_{\le r}\big((af) \star_\theta h\big)$ and $a\,\pi_{\le r}\{f,h\}_\theta = \pi_{\le r}\{af,h\}_\theta$; hence $\mathcal{G}^{(\le K)}$ is closed under real multiples and, by (36),

$$\begin{aligned} \exp\big(s\,\mathcal{L}_{g,\le r}\big) = \exp\big(\epsilon\,\mathcal{L}_{(s/\epsilon)g,\le r}\big) = \Phi_{\epsilon,(s/\epsilon)g}, &\qquad s \in \mathbb{R},\ \epsilon > 0, \\ \Phi_{\epsilon,g}^{-1} = \exp\big(-\epsilon\,\mathcal{L}_{g,\le r}\big) = \Phi_{\epsilon,-g}. & \end{aligned} \quad (42)$$

By (42), every element of $\mathscr{T}_{r,K}$ is a finite composition $\Phi_{\epsilon,g_j}\circ\cdots\circ\Phi_{\epsilon,g_1}$ of DAG-indexed drivers, and the family in (34), closed under composition and inversion, forms the finite-dimensional Lie group $\mathscr{T}_{r,K}$, completing the proof. □

Proposition A.4 (Planar-binary-DAG correspondence, quotient form)**.** *For every $k \ge 0$ there exists a bijection*

$$\Phi_k : \mathcal{G}^{(k)} \longrightarrow \mathcal{T}^{(k)}$$

between the depth-k Neural–Poisson generators modulo the relations of the projected Poisson algebra and the planar binary DAGs of height k modulo the corresponding local moves. Consequently $\mathcal{G}^{(\le K)} \cong \bigcup_{h=0}^{K} \mathcal{T}^{(h)}$ for every $K \ge 0$.

Proof. Let $\mathcal{G}^{(0)} = \mathcal{G}_\theta = \{g_\theta^{(1)},\dots,g_\theta^{(m)}\}$. For $k \ge 1$ let $\widetilde{\mathcal{G}}^{(k)}$ be the set of formal binary expressions obtained from ordered pairs $(f,g) \in \widetilde{\mathcal{G}}^{(k-1)} \times \widetilde{\mathcal{G}}^{(k-1)}$ by the two constructors

$$(f,g) \mapsto \pi_{\le r}(f \star_\theta g), \qquad (f,g) \mapsto \pi_{\le r}(\{f,g\}_\theta). \quad (43)$$

Let $\sim$ be the smallest congruence on $\bigcup_{h\le k} \widetilde{\mathcal{G}}^{(h)}$ generated by the following relations: commutativity of $\star_\theta$, antisymmetry of the bracket, the Jacobi identity, and the idempotence of the truncation $\pi_{\le r}$. Let $\widetilde{\mathcal{T}}^{(k)}$ be the set of planar binary DAGs of height k with leaves labelled by $\mathcal{G}_\theta$ and internal nodes

labelled by $\star_\theta$ or $\{\cdot,\cdot\}_\theta$ with ordered children. Let $\equiv$ be the smallest congruence on $\bigcup_{h\le k}\widetilde{\mathcal{T}}^{(h)}$ generated by the corresponding local moves: swapping the two children at a $\star_\theta$-node, swapping the two children at a bracket node together with the induced global sign flip, the Jacobi rewrite on depth two bracket patterns, and erasing duplicate truncation tags. The two quotients in the statement are

$$\mathcal{G}^{(k)} := \widetilde{\mathcal{G}}^{(k)}/\sim, \qquad \mathcal{T}^{(k)} := \widetilde{\mathcal{T}}^{(k)}/\equiv . \tag{44}$$

We construct mutually inverse maps between the quotients (44) by structural recursion on the depth. For $k = 0$ let $\Phi_0(g_\theta^{(i)})$ be the one-node DAG labelled by $g_\theta^{(i)}$. For $k \ge 1$ and a representative $E \in \widetilde{\mathcal{G}}^{(k)}$ of $[E] \in \mathcal{G}^{(k)}$, built by one of the constructors (43), define

$$\widetilde{\Phi}_k(E) := \begin{cases} \widetilde{\Phi}_{k-1}(f) \;\; \overset{\star_\theta}{\frown} \;\; \widetilde{\Phi}_{k-1}(g) & \text{if } E = \pi_{\le r}(f \star_\theta g), \\ \qquad\qquad\qquad\qquad\qquad\qquad & \qquad \Phi_k([E]) := \big[\widetilde{\Phi}_k(E)\big] \in \mathcal{T}^{(k)}. \\ \widetilde{\Phi}_{k-1}(f) \;\; \overset{\{\cdot,\cdot\}_\theta}{\frown} \;\; \widetilde{\Phi}_{k-1}(g) & \text{if } E = \pi_{\le r}(\{f, g\}_\theta), \end{cases} \tag{45}$$

If $E \sim E'$, the two expressions differ by a finite sequence of generators of $\sim$, and (45) sends each generator to a generator of $\equiv$:

$$\begin{aligned} \pi_{\le r}(f \star_\theta g) \sim \pi_{\le r}(g \star_\theta f) &\;\longmapsto\; \text{swap of the children at a } \star_\theta \text{ -node}, \\ \pi_{\le r}\{f, g\}_\theta \sim -\pi_{\le r}\{g, f\}_\theta &\;\longmapsto\; \text{swap at a bracket node with sign flip}, \\ \{f,\{g,h\}_\theta\}_\theta + \{g,\{h,f\}_\theta\}_\theta + \{h,\{f,g\}_\theta\}_\theta \sim 0 &\;\longmapsto\; \text{Jacobi rewrite (depth two)}, \\ \pi_{\le r} \circ \pi_{\le r} = \pi_{\le r} &\;\longmapsto\; \text{erasure of a duplicate tag}. \end{aligned} \tag{46}$$

Hence $\widetilde{\Phi}_k(E) \equiv \widetilde{\Phi}_k(E')$ and Φ_k is well defined. In the reverse direction, let $\Psi_0 := \Phi_0^{-1}$ on leaves, and for $k \ge 1$ and a representative $T \in \widetilde{\mathcal{T}}^{(k)}$ of $[T] \in \mathcal{T}^{(k)}$ with children T_L, T_R define

$$\begin{aligned} \widetilde{\Psi}_k\Big(\; T_L \;\; \overset{\star_\theta}{\frown} \;\; T_R \; \Big) &:= \pi_{\le r}\big(\widetilde{\Psi}_{k-1}(T_L) \star_\theta \widetilde{\Psi}_{k-1}(T_R)\big), \\ \widetilde{\Psi}_k\Big(\; T_L \;\; \overset{\{\cdot,\cdot\}_\theta}{\frown} \;\; T_R \; \Big) &:= \pi_{\le r}\big(\{\widetilde{\Psi}_{k-1}(T_L), \widetilde{\Psi}_{k-1}(T_R)\}_\theta\big), \\ \Psi_k([T]) &:= \big[\widetilde{\Psi}_k(T)\big] \in \mathcal{G}^{(k)}. \end{aligned} \tag{47}$$

Each local move generating $\equiv$ is sent by (47) to the corresponding generator of $\sim$ in (46), read from right to left, or to the idempotence of $\pi_{\le r}$. Hence $T \equiv T'$ implies $\widetilde{\Psi}_k(T) \sim \widetilde{\Psi}_k(T')$ and Ψ_k is well defined.

We show $\Psi_k \circ \Phi_k = \mathrm{Id}$ and $\Phi_k \circ \Psi_k = \mathrm{Id}$ by induction on k; the case $k = 0$ holds by definition. Assume $\widetilde{\Psi}_{k-1}(\widetilde{\Phi}_{k-1}(f)) \sim f$ and $\widetilde{\Phi}_{k-1}(\widetilde{\Psi}_{k-1}(T)) \equiv T$ at depth $k-1$. For $[E] \in \mathcal{G}^{(k)}$ with representative $E = \pi_{\le r}(f \star_\theta g)$, the definitions (45)–(47), the induction hypothesis, and the congruence property of $\sim$ give

$$\begin{aligned} \Psi_k\big(\Phi_k([E])\big) &= \Big[\widetilde{\Psi}_k\big(\widetilde{\Phi}_k(E)\big)\Big] = \Big[\, \pi_{\le r}\Big(\widetilde{\Psi}_{k-1}\big(\widetilde{\Phi}_{k-1}(f)\big) \star_\theta \widetilde{\Psi}_{k-1}\big(\widetilde{\Phi}_{k-1}(g)\big)\Big) \,\Big] \\ &= \big[\, \pi_{\le r}(f \star_\theta g) \,\big] = [E], \end{aligned} \tag{48}$$

and the bracket case is identical. Conversely, for $[T] \in \mathcal{T}^{(k)}$ with root operation $\odot \in \{\star_\theta, \{\cdot,\cdot\}_\theta\}$ and children T_L, T_R, the induction hypothesis and the congruence property of $\equiv$ give

$$\Phi_k\big(\Psi_k([T])\big) = \Big[\;\; \widetilde{\Phi}_{k-1}(\widetilde{\Psi}_{k-1}(T_L)) \;\; \overset{\odot}{\frown} \;\; \widetilde{\Phi}_{k-1}(\widetilde{\Psi}_{k-1}(T_R)) \;\; \Big] = \Big[\; T_L \;\; \overset{\odot}{\frown} \;\; T_R \; \Big] = [T]. \tag{49}$$

By (48) and (49), Φ_k and Ψ_k are inverse bijections. The depth classes are pairwise disjoint, and the maps Φ_h, $0 \le h \le K$, assemble into the bijection

$$\mathcal{G}^{(\le K)} = \bigcup_{h=0}^{K} \mathcal{G}^{(h)} \;\xrightarrow{\;\Phi_h \text{ on } \mathcal{G}^{(h)}\;}\; \bigcup_{h=0}^{K} \mathcal{T}^{(h)}, \tag{50}$$

which proves $\mathcal{G}^{(\le K)} \cong \bigcup_{h=0}^{K} \mathcal{T}^{(h)}$. □

A.3 Generalization

In preparation for the proof of generalization bound, we first recall the essential definitions and properties of Hamiltonian vector fields, the polynomial truncation operator $\pi_{\leq r}$, and the ambient Lie algebra $\mathfrak{X}(\mathcal{M})$ that will be used in our subsequent argument.

Lie Subalgebra Structure of $\mathfrak{t}_{r,K}$. Let $\mathfrak{X}(\mathcal{M})$ denote the Lie algebra of smooth vector fields on the symplectic manifold $(\mathcal{M}, \omega)$, equipped with the standard Lie bracket $[X, Y] = X \circ Y - Y \circ X$. For any Hamiltonian function $f \in C^\infty(\mathcal{M})$, the associated Hamiltonian vector field L_f is defined by $\iota_{L_f}\omega = df$. Truncating to polynomial Hamiltonians of degree at most r, write $L_{f,\leq r} := \pi_{\leq r}(L_f)$, where $\pi_{\leq r}$ projects onto the space of polynomial vector fields of degree $\leq r$. We define the set of Lie transformation as follows:

$$\mathfrak{t}_{r,K} := \operatorname{span}\big\{ L_{g,\leq r} \;\big|\; g \in G^{(\leq K)} \big\} \subset \mathfrak{X}(\mathcal{M}).$$

Note that the Poisson bracket of any two truncated polynomials in $\mathcal{G}^{(\leq K)}$ remains in $\mathcal{G}^{(\leq K)}$, and commutators of Hamiltonian vector fields satisfy the following:

$$[L_{g_1,\leq r}, L_{g_2,\leq r}] = L_{\{g_1,g_2\},\leq r}, \quad g_1, g_2 \in \mathcal{G}^{(\leq K)},$$

Thus, it follows that $\mathfrak{t}_{r,K}$ is closed under the Lie bracket, showing that $\mathfrak{t}_{r,K}$ is a finite-dimensional Lie *subalgebra* of $\mathfrak{X}(\mathcal{M})$.

Compatible Almost–Complex Structure. Let $(\mathcal{M}, \omega)$ be a $2n$–dimensional symplectic manifold. An *almost–complex structure* on $\mathcal{M}$ is a bundle map $J\colon T\mathcal{M} \longrightarrow T\mathcal{M}$ satisfying $J^2 = -\mathrm{Id}$. We say that J is *compatible* with the symplectic form ω if, for all tangent vectors $u, v \in T\mathcal{M}$,

$$\omega(Ju, Jv) = \omega(u, v), \qquad \omega(u, Ju) > 0 \quad (\text{for } u \neq 0) \tag{51}$$

In particular, these conditions guarantee that the bilinear form $g(u, v) := \omega(u, Jv)$ is symmetric and positive–definite, hence defines a Riemannian metric on M. Together, (ω, J, g) form an *almost–Kähler* structure, which underlies the L^2–inner product

$$\big\langle X, Y \big\rangle = \int_M \omega\big(X(z), JY(z)\big) d\mu_\omega(z) \quad \forall\ X, Y \in \mathfrak{X}(\mathcal{M}), \tag{52}$$

and ensures that Hamiltonian flows preserve both the symplectic form and the associated phase-space volume $\mu_\omega = \omega^n/n!$ [McDuff and Salamon, 1998, da Silva, 2008].

Left-Invariant Geometry of $\mathscr{T}_{r,K}$. We extend the proposed inner product to a Riemannian metric on the Lie group $\mathscr{T}_{r,K}$ by left translations. For any fixed Lie element $g \in \mathscr{T}_{r,K}$ and tangent vectors $X, Y \in T_g\mathscr{T}_{r,K} \subset \mathfrak{X}(\mathcal{M})$, we can rewrite the inner product by realizing the sub-algebra structure:

$$\begin{aligned}
\big\langle X_g, Y_g \big\rangle_g &= \big\langle \underbrace{d(L_{g^{-1}})_g[X_g]}_{=:U}, \underbrace{d(L_{g^{-1}})_g[Y_g]}_{=:V} \big\rangle \quad (\text{by definition of the left-invariant metric}) && (53)\\
&= \int_M \omega\big(U(z), JV(z)\big) \mathrm{d}\mu_\omega(z) \quad (\text{by the global } L^2\text{–inner product on } \mathfrak{X}(M)) && (54)\\
&= \int_M \omega\big(d(L_{g^{-1}})_g[X_g](z), Jd(L_{g^{-1}})_g[Y_g](z)\big) \mathrm{d}\mu_\omega(z). && (55)
\end{aligned}$$

where $L_{g^{-1}}\colon h \mapsto g^{-1}h$ is left multiplication. By construction, this metric is *left-invariant*, since left-translation by any group element preserves the inner product on each tangent space. Denote by d_G the geodesic distance induced by this Riemannian metric:

$$d_G(g, h) = \inf_{\substack{\gamma(0)=g \\ \gamma(1)=h}} \int_0^1 \big\|\dot\gamma(t)\big\|_{\gamma(t)} dt,$$

where the infimum is over all smooth curves $\gamma\colon [0, 1] \to \mathscr{T}_{r,K}$ connecting g to h. Given definition of geodesic, the *exponential map* is simply given as $\exp\colon \mathfrak{t}_{r,K} \longrightarrow \mathscr{T}_{r,K}$, where the map is defined by sending a Lie algebra element X to the time-one point $\gamma_X(1)$ of the unique geodesic γ_X satisfying $\gamma_X(0) = e$ and $\dot\gamma_X(0) = X$.

Covering Number of Neural-Poisson Lie Walk. Exploiting the left-invariant Riemannian structure of $\mathscr{T}_{r,K}$ and the properties of the geodesic exponential map introduced above, we now proceed to

derive an upper bound on the covering number of the elementary-flow set $\mathscr{S}_h$ with respect to the distance d_G. Let $\mathcal{W}^{\theta}_{r,K,M,h}$ be the Neural–Poisson Lie–group walk class. For a maximal step size $h > 0$ define the *elementary-flow set*

$$\mathscr{S}_h := \Big\{ \Phi_{\varepsilon,g} = \exp\big(\varepsilon L_{g,\le r}\big) \;\Big|\; g \in G^{(\le K)}, 0 < \varepsilon \le h \Big\} \subset \mathscr{T}_{r,K}, \tag{56}$$

so every element is the ε-time Lie–exponential generated by a depth $\le K$ polynomial Hamiltonian, truncated at degree r. The left-invariant Riemannian distance d_G on $\mathscr{T}_{r,K}$ restricts to a metric on $\mathscr{S}_h$. We then define the *covering number* of this elementary-flow set by

$$N(\delta, \mathscr{S}_h, d_G) := \min\Big\{ N \in \mathbb{N} \;\Big|\; \exists\, g_1, \dots, g_N \in \mathscr{S}_h, \quad \mathscr{S}_h \subset \bigcup_{i=1}^{N} B_{d_G}\big(g_i, \varepsilon\big) \Big\},$$

which captures the minimal number of geodesic balls of radius ε needed to cover $\mathscr{S}_h$ and thus quantifies its metric complexity at scale δ. Next, for each tolerance δ_j, let $\mathcal{C}_{\delta_j} \subset \mathscr{S}_h$ be any δ_j-cover (or δ_j-net) satisfying

$$\big|\mathcal{C}_{\delta_j}\big| = N\big(\delta_j, \mathscr{S}_h, d_G\big), \qquad \mathscr{S}_h \subset \bigcup_{\Phi \in \mathcal{C}_{\delta_j}} B_{d_G}(\Phi, \delta_j),$$

where the geodesic ball B_{d_G} centered at Φ in $\mathscr{S}_h$ is defined by

$$B_{d_G}(\Phi, \delta_j) := \big\{ \Psi \in \mathscr{S}_h \mid d_G(\Phi, \Psi) < \delta_j \big\}.$$

In other words, $\mathcal{C}_{\delta_j}$ is the set of centers of open balls of radius δ_j that cover $\mathscr{S}_h$, and its cardinality exactly equals the covering number.

Proposition A.5 (Compositional entropy bound for Lie–group walks)**.** *Let $M \in \mathbb{N}$. For a sequence of non–negative tolerances $\delta_1, \dots, \delta_M$ satisfying $\sum_{j=1}^{M} \delta_j \le \varepsilon$ define the Cartesian product of step covers*

$$\mathcal{C}(\delta_1, \dots, \delta_M) := \mathcal{C}_{\delta_M} \times \cdots \times \mathcal{C}_{\delta_1}, \quad \mathcal{C}_{\delta_j} \subset \mathscr{S}_h, \quad |\mathcal{C}_{\delta_j}| = N(\delta_j, \mathscr{S}_h, d_G).$$

For every choice $(\widehat{\Phi}_M, \dots, \widehat{\Phi}_1) \in \mathcal{C}(\delta_1, \dots, \delta_M)$ form the composite $\widehat{W} := \widehat{\Phi}_M \circ \cdots \circ \widehat{\Phi}_1$. Then the collection $\{\widehat{W}\}$ is an ε-net of the Lie–group walk class $\mathcal{W}^{\theta}_{r,K,M,h} = \{\Phi_{\varepsilon_M, g_M} \circ \cdots \circ \Phi_{\varepsilon_1, g_1} \mid (\varepsilon_j, g_j)$ admissible$\}$, and the covering number satisfies the purely factorised *bound*

$$N\big(\varepsilon, \mathcal{W}^{\theta}_{r,K,M,h}, d_G\big) \le \prod_{j=1}^{M} N\big(\delta_j, \mathscr{S}_h, d_G\big).$$

In particular, choosing the uniform allocation $\delta_1 = \cdots = \delta_M = \varepsilon/M$ yields

$$N\big(\varepsilon, \mathcal{W}^{\theta}_{r,K,M,h}, d_G\big) \le \Big[N\big(\varepsilon/M, \mathscr{S}_h, d_G\big)\Big]^{M}.$$

Proof. Fix an admissible Lie group walk and, for every index $j \in \{1, \dots, M\}$, a proxy from the δ_j-cover,

$$W_{\mathbf{g}}(M) := \Phi_{\varepsilon_M, g_M} \circ \cdots \circ \Phi_{\varepsilon_1, g_1} \in \mathcal{W}^{\theta}_{r,K,M,h}, \qquad \widehat{\Phi}_j \in \mathcal{C}_{\delta_j}, \quad d_G\big(\Phi_{\varepsilon_j, g_j}, \widehat{\Phi}_j\big) < \delta_j. \tag{57}$$

Define the mixed partial compositions

$$W^{(j)} := \widehat{\Phi}_M \circ \cdots \circ \widehat{\Phi}_{j+1} \circ \Phi_{\varepsilon_j, g_j} \circ \cdots \circ \Phi_{\varepsilon_1, g_1}, \qquad j = 0, \dots, M, \tag{58}$$

with $W^{(0)} = W_{\mathbf{g}}(M)$ and $W^{(M)} = \widehat{\Phi}_M \circ \cdots \circ \widehat{\Phi}_1 =: \widehat{W}$. Consecutive terms of (58) differ only in the j-th factor; with the suffix $S := \Phi_{\varepsilon_{j-1}, g_{j-1}} \circ \cdots \circ \Phi_{\varepsilon_1, g_1}$ (and $S = \mathbf{Id}$ for $j = 1$), the triangle inequality gives

$$W^{(j-1)} = \widehat{\Phi}_M \circ \cdots \circ \widehat{\Phi}_{j+1} \circ \underbrace{\Phi_{\varepsilon_j, g_j}}_{\text{varying}} \circ S, \qquad W^{(j)} = \widehat{\Phi}_M \circ \cdots \circ \widehat{\Phi}_{j+1} \circ \underbrace{\widehat{\Phi}_j}_{\text{proxy}} \circ S,$$

$$d_G\big(W_{\mathbf{g}}(M), \widehat{W}\big) = d_G\big(W^{(0)}, W^{(M)}\big) \le \sum_{j=1}^{M} d_G\big(W^{(j-1)}, W^{(j)}\big). \tag{59}$$

Left translation is an isometry of d_G. Right translation by S maps a curve γ joining $\widehat{\Phi}_j$ to Φ_{ε_j,g_j} onto the curve $\gamma \circ S$ joining $\widehat{\Phi}_j \circ S$ to $\Phi_{\varepsilon_j,g_j} \circ S$, whose left-trivialized velocity is the adjoint image of that of γ:

$$\begin{aligned}\left\| \tfrac{d}{dt}\big(\gamma(t) \circ S\big) \right\|_{\gamma(t)\circ S} &= \left\| (\gamma(t) \circ S)^{-1} \circ \dot\gamma(t) \circ S \right\|_e = \left\| \mathrm{Ad}_{S^{-1}}\big(\gamma(t)^{-1} \circ \dot\gamma(t)\big) \right\|_e \\ &\le \left\| \mathrm{Ad}_{S^{-1}} \right\|_{\mathrm{op}} \|\dot\gamma(t)\|_{\gamma(t)}. \end{aligned} \tag{60}$$

Integrating (60) over $t \in [0,1]$, taking the infimum over γ, and using left-invariance for the common prefix and (57) for the proxy,

$$d_G\big(W^{(j-1)}, W^{(j)}\big) = d_G\big(\Phi_{\varepsilon_j,g_j} \circ S,\ \widehat{\Phi}_j \circ S\big) \le \left\| \mathrm{Ad}_{S^{-1}} \right\|_{\mathrm{op}} d_G\big(\Phi_{\varepsilon_j,g_j},\ \widehat{\Phi}_j\big) < \left\| \mathrm{Ad}_{S^{-1}} \right\|_{\mathrm{op}} \delta_j. \tag{61}$$

We equip $\mathscr{T}_{r,K}$ with the left-invariant metric whose inner product at the identity is invariant under the adjoint action of the walk prefixes, i.e. a bi-invariant metric along the walk, for which $\| \mathrm{Ad}_{S^{-1}} \|_{\mathrm{op}} = 1$ in (61); for a general left-invariant metric the factor $\sup_S \| \mathrm{Ad}_{S^{-1}} \|_{\mathrm{op}}$ multiplies all radii below. Inserting (61) into (59),

$$d_G\big(W_{\mathbf{g}}(M), \widehat{W}\big) \le \sum_{j=1}^{M} d_G\big(W^{(j-1)}, W^{(j)}\big) < \sum_{j=1}^{M} \delta_j \le \varepsilon, \tag{62}$$

hence every walk lies in the ε-ball centred at some composite $\widehat{W}$, and $\{\widehat{W}\}$ is an ε-net of $\mathcal{W}^{\theta}_{r,K,M,h}$. The composites in (62) are indexed by the tuples $(\widehat{\Phi}_M, \dots, \widehat{\Phi}_1) \in \mathcal{C}(\delta_1, \dots, \delta_M)$, hence

$$N\big(\varepsilon, \mathcal{W}^{\theta}_{r,K,M,h}, d_G\big) \le \big|\mathcal{C}(\delta_1, \dots, \delta_M)\big| = \prod_{j=1}^{M} |\mathcal{C}_{\delta_j}| = \prod_{j=1}^{M} N\big(\delta_j, \mathscr{S}_h, d_G\big), \tag{63}$$

and the uniform allocation $\delta_1 = \cdots = \delta_M = \varepsilon/M$ in (63) gives

$$N\big(\varepsilon, \mathcal{W}^{\theta}_{r,K,M,h}, d_G\big) \le \prod_{j=1}^{M} N\big(\tfrac{\varepsilon}{M}, \mathscr{S}_h, d_G\big) = \Big[N\big(\tfrac{\varepsilon}{M}, \mathscr{S}_h, d_G\big)\Big]^M.$$

□

We now derive a covering number bound for the elementary flow set $\mathscr{S}_h$ under the geodesic metric d_G. Recall that

$$\mathscr{S}_h := \{\Phi_{\varepsilon,g} = \exp(\varepsilon L_{g,\le r}) | g \in \mathcal{G}(\le K), 0 < \varepsilon \le h\} \subset \mathcal{T}_{r,K},$$

where $\mathcal{G}(\le K)$ denotes the finite set of depth-$\le K$ polynomial Hamiltonians (with at most D free parameters), and $L_{g,\le r}$ is the truncated Lie operator. The set $\mathscr{S}_h$ thus forms a compact, finite-dimensional submanifold of the Lie group $\mathcal{T}_{r,K}$, parameterized by (ε, g). Next, we estimate its covering number under the geodesic distance d_G induced by any left-invariant Riemannian metric.

Proposition A.6 (Volume–packing covering bound)**.** *Let (M, g) be a compact d-dimensional Riemannian manifold with geodesic metric d_g, total volume $\mathrm{Vol}(M)$, and injectivity radius $\mathrm{Inj}(M) > 0$. For every radius $0 < \delta \le \mathrm{Inj}(M)/2$ the δ-covering number satisfies*

$$N\big(\delta, M, d_g\big) \le C_M \delta^{-d}, \qquad C_M = \frac{2^d \mathrm{Vol}(M)}{c_M},$$

where the uniform lower-volume constant is defined by

$$c_M := \inf_{x \in M} \inf_{0 < r \le \mathrm{Inj}(M)/2} \frac{\mathrm{Vol}\big(B_g(x,r)\big)}{r^d} > 0.$$

Proof. Let $\mathrm{Inj}(M)$ denote the injectivity radius of (M, g), and define the volume ratio of geodesic balls,

$$\mathrm{Inj}(M) = \inf_{x \in M} \sup\big\{r > 0 : \exp_x \colon B_{T_xM}(0, r) \to M \text{ is a diffeomorphism}\big\},$$
$$f(x, r) := \frac{\mathrm{Vol}\big(B_g(x, r)\big)}{r^d}, \qquad 0 < r \le \tfrac{\mathrm{Inj}(M)}{2}. \tag{64}$$

For $r \le \mathrm{Inj}(M)/2$ the ball $B_g(x, r)$ lies inside a normal-coordinate chart around x, in which the metric agrees with the Euclidean one up to second-order terms; with $\omega_d := \mathrm{Vol}_{\mathbb{R}^d}\big(B_{\mathbb{R}^d}(0, 1)\big) = \pi^{d/2}/\Gamma\big(\frac{d}{2} + 1\big)$ this gives

$$\lim_{r \to 0} \frac{\mathrm{Vol}\big(B_g(x, r)\big)}{\mathrm{Vol}_{\mathbb{R}^d}\big(B(0, r)\big)} = 1 \quad \Longrightarrow \quad \lim_{r \to 0} f(x, r) = \lim_{r \to 0} \frac{\mathrm{Vol}\big(B_g(x, r)\big)}{\omega_d\, r^d}\, \omega_d = \omega_d. \tag{65}$$

By (65), setting $f(x, 0) := \omega_d$ extends the ratio (64) to a continuous positive function on the compact set $M \times [0, \mathrm{Inj}(M)/2]$, which therefore attains a positive minimum:

$$c_M := \min_{(x,r) \in M \times [0, \mathrm{Inj}(M)/2]} f(x, r) > 0 \quad \Longrightarrow \quad \mathrm{Vol}\big(B_g(x, r)\big) \ge c_M\, r^d,$$
$$x \in M, \;\; 0 \le r \le \tfrac{\mathrm{Inj}(M)}{2}. \tag{66}$$

Fix $0 < \delta \le \mathrm{Inj}(M)/2$ and let $\{x_1, \dots, x_N\} \subset M$ be a maximal δ-separated set, $d_g(x_i, x_j) \ge \delta$ for $i \ne j$. Maximality means that every $x \in M$ satisfies $d_g(x, x_i) < \delta$ for some i, hence the centres form a δ-cover and

$$M = \bigcup_{i=1}^{N} B_g(x_i, \delta), \qquad N(\delta, M, d_g) \le N. \tag{67}$$

The half-radius balls are pairwise disjoint: a point $z \in B_g(x_i, \frac{\delta}{2}) \cap B_g(x_j, \frac{\delta}{2})$ with $i \ne j$ would give

$$\delta \le d_g(x_i, x_j) \le d_g(x_i, z) + d_g(z, x_j) < \tfrac{\delta}{2} + \tfrac{\delta}{2} = \delta, \tag{68}$$

a contradiction. Disjointness and (66) with $r = \delta/2 \le \mathrm{Inj}(M)/2$ yield

$$\mathrm{Vol}(M) \ge \mathrm{Vol}\Big(\bigcup_{i=1}^{N} B_g\big(x_i, \tfrac{\delta}{2}\big)\Big) = \sum_{i=1}^{N} \mathrm{Vol}\big(B_g(x_i, \tfrac{\delta}{2})\big) \ge N\, c_M \Big(\frac{\delta}{2}\Big)^d = N\, \frac{c_M\, \delta^d}{2^d}. \tag{69}$$

Solving (69) for N and combining with (67),

$$N(\delta, M, d_g) \le N \le \frac{2^d\, \mathrm{Vol}(M)}{c_M}\, \delta^{-d} = C_M\, \delta^{-d}, \qquad C_M = \frac{2^d\, \mathrm{Vol}(M)}{c_M}.$$

□

Proposition A.7 (Covering number of the elementary–flow Lie subgroup)**.** *Equip $\mathscr{T}_{r,K}$ with the left–invariant geodesic metric d_G, and let $d_{\mathscr{S}}$ be the dimension of the elementary-flow set $\mathscr{S}_h$ with step ceiling $h > 0$. Then for every radius $0 < \delta \le \mathrm{Inj}(\mathscr{S}_h)/2$,*

$$N\big(\delta, \mathscr{S}_h, d_G\big) \le C_{\mathscr{S}}\, \delta^{-d_{\mathscr{S}}}, \qquad C_{\mathscr{S}} = \frac{2^{d_{\mathscr{S}}}\, \mathrm{Vol}(\mathscr{S}_h)}{c_{\mathscr{S}}},$$

where $c_{\mathscr{S}} > 0$ is the uniform lower-volume constant of $\mathscr{S}_h$.

Proof. For a fixed step ceiling $h > 0$, the elementary-flow set and the uniform lower–volume constant in the statement are

$$\mathscr{S}_h = \Big\{\Phi_{\varepsilon, g} = \exp\big(\varepsilon L_{g, \le r}\big) \;\Big|\; g \in \mathcal{G}^{(\le K)},\; 0 < \varepsilon \le h\Big\} \subset \mathscr{T}_{r,K}, \tag{70}$$

$$c_{\mathscr{S}} := \inf_{x \in \mathscr{S}_h} \inf_{0 < r \le \mathrm{Inj}(\mathscr{S}_h)/2} \frac{\mathrm{Vol}\big(B_G(x, r)\big)}{r^{d_{\mathscr{S}}}} > 0, \tag{71}$$

where $d_{\mathscr{S}} := \dim \mathscr{S}_h$, $\mathrm{Vol}(\mathscr{S}_h)$ is the total Riemannian volume, $\mathrm{Inj}(\mathscr{S}_h) > 0$ is the injectivity radius with respect to d_G, and $B_G(x, r)$ denotes the geodesic ball of radius r centered at x in the metric d_G. The set (70) is a compact $d_{\mathscr{S}}$-dimensional submanifold of the Lie group, endowed with the restriction of the left-invariant Riemannian metric, hence a compact Riemannian manifold with geodesic metric d_G. Proposition A.6 applies under the identification

$$(M,\ d_g,\ d,\ \mathrm{Vol}(M),\ \mathrm{Inj}(M),\ c_M) \longmapsto (\mathscr{S}_h,\ d_G,\ d_{\mathscr{S}},\ \mathrm{Vol}(\mathscr{S}_h),\ \mathrm{Inj}(\mathscr{S}_h),\ c_{\mathscr{S}}), \tag{72}$$

with c_M replaced by (71), and gives, for every $0 < \delta \le \mathrm{Inj}(\mathscr{S}_h)/2$,

$$N\big(\delta, \mathscr{S}_h, d_G\big) \le \frac{2^{d_{\mathscr{S}}}\,\mathrm{Vol}(\mathscr{S}_h)}{c_{\mathscr{S}}}\ \delta^{-d_{\mathscr{S}}} = C_{\mathscr{S}}\,\delta^{-d_{\mathscr{S}}}.$$

□

Proposition A.8. *Let $\mathscr{S}_h^{\Theta}$ be the elementary-flow set parameterized by the step size $\varepsilon \in (0, h]$ and the coefficients $c \in \Omega \subset \mathbb{R}^d$ of the polynomial Hamiltonian $g(c)$, and assume that the left-invariant metric of $\mathscr{T}_{r,K}$ is block-diagonal in the coordinates (ε, c). Then the Riemannian volume of $\mathscr{S}_h^{\Theta}$ grows at most linearly in the cardinality of the Poisson DAG:*

$$\mathrm{Vol}\big(\mathscr{S}_h^{\Theta}\big) \le h\,|G^{(\le K)}|\,\widehat{V}\,\widehat{\mathcal{J}},$$

where $\widehat{V}$ and $\widehat{\mathcal{J}}$ are the largest Lebesgue volume and the largest Jacobian factor of the coefficient tiles.

Proof. Here $c = (c_1, \dots, c_d)$ are the continuous coefficients of the polynomial Hamiltonian $g(c)$, Ω is the domain of allowable coefficients, and

$$\mathscr{S}_h^{\Theta} = \big\{\Phi_{\varepsilon, g(c)} = \exp(\varepsilon L_{g(c),\le r}) \;\big|\; 0 < \varepsilon \le h,\ c \in \Omega \subset \mathbb{R}^d\big\}. \tag{73}$$

The block-diagonal assumption states that the left-invariant Riemannian metric $g^{\mathscr{T}}$ on $\mathscr{T}_{r,K}$ pulls back under the mapping ϕ to a block-diagonal metric of the form

$$\begin{aligned} \phi\colon (0,h] \times \Omega \longrightarrow \mathscr{S}_h^{\Theta}, \qquad & \phi(\varepsilon, c) = \exp\big(\varepsilon L_{g(c),\le r}\big), \\ \phi^* g^{\mathscr{T}} = d\varepsilon^2 \oplus g_{\mathrm{fiber}}(c), \qquad & g_{\mathrm{fiber}}(c) \equiv \Sigma_\theta \in \mathbb{R}^{d\times d}. \end{aligned} \tag{74}$$

The map ϕ is a smooth bijection from $(0, h] \times \Omega$ onto (73) and provides global coordinates (ε, c). Left-translating the coordinate derivatives back to the identity e gives the blocks of the pulled-back metric. In the longitudinal direction, $\partial_\varepsilon \exp(\varepsilon X) = \exp(\varepsilon X) \circ X$ yields

$$\begin{aligned} d\big(L_{\phi(\varepsilon,c)^{-1}}\big)\,\partial_\varepsilon \phi(\varepsilon, c) &= \exp\big(-\varepsilon L_{g(c),\le r}\big) \circ \exp\big(\varepsilon L_{g(c),\le r}\big) \circ L_{g(c),\le r} = L_{g(c),\le r}, \\ A(c) := \big(\phi^* g^{\mathscr{T}}\big)_{\varepsilon\varepsilon} &= \big\langle L_{g(c),\le r}, L_{g(c),\le r}\big\rangle_e, \end{aligned} \tag{75}$$

and the normalization $\langle L_{g(c),\le r}, L_{g(c),\le r}\rangle_e = 1$ of the generators gives the block $d\varepsilon^2$ in (74). In the coefficient directions $v, w \in T_c\Omega$, the fiber block is

$$g_{\mathrm{fiber}}(c)[v, w] := \big\langle d\big(L_{\phi^{-1}}\big)\partial_v \phi,\ d\big(L_{\phi^{-1}}\big)\partial_w \phi\big\rangle_e = \big(\Sigma_\theta(c)\big)_{vw}, \tag{76}$$

and (74) states that the mixed (ε, c) entries vanish. Hence, by (75) and (76), in coordinates (ε, c),

$$\phi^* g^{\mathscr{T}} = \begin{pmatrix} A(c) & 0 \\ 0 & \Sigma_\theta(c) \end{pmatrix}, \qquad \det\big(\phi^* g^{\mathscr{T}}\big) = A(c)\,\det \Sigma_\theta(c), \tag{77}$$

where we keep the dependence on c explicit; the statement corresponds to $A \equiv 1$ and $\Sigma_\theta(c) \equiv \Sigma_\theta$. The Riemannian volume form is $d\mu_{\mathscr{S}_h} = \sqrt{\det(\phi^* g^{\mathscr{T}})}\,d\varepsilon\,dc$, and by (77) the integrand is independent of ε:

$$\begin{aligned} \mathrm{Vol}\big(\mathscr{S}_h^{\Theta}\big) &= \int_\Omega \int_0^h \sqrt{\det\big(\phi^* g^{\mathscr{T}}\big)(\varepsilon, c)}\ d\varepsilon\,dc = \int_\Omega \int_0^h \sqrt{A(c)\det \Sigma_\theta(c)}\ d\varepsilon\,dc \\ &= h \int_\Omega \sqrt{A(c)\det \Sigma_\theta(c)}\ dc. \end{aligned} \tag{78}$$

Let $\Psi\colon \Omega \to G^{(\le K)}$, $c \mapsto g(c)$, be surjective, and define for each $g \in G^{(\le K)}$ the coefficient tile, its Lebesgue measure, and its Jacobian factor,

$$T_g := \Psi^{-1}(g) = \{c \in \Omega \mid \Psi(c) = g\}, \qquad \mathrm{Vol}_g := \lambda(T_g), \qquad \widehat{\mathcal{J}}_g := \sup_{c \in T_g} \sqrt{A(c) \det \Sigma_\theta(c)}. \tag{79}$$

The tiles (79) partition $\Omega = \bigsqcup_g T_g$; with $\widehat{V} := \max_g \mathrm{Vol}_g$ and $\widehat{\mathcal{J}} := \max_g \widehat{\mathcal{J}}_g$, (78) gives

$$\begin{aligned} \mathrm{Vol}\big(\mathscr{S}_h^\Theta\big) &= h \sum_{g \in G^{(\le K)}} \int_{T_g} \sqrt{A(c) \det \Sigma_\theta(c)}\, dc \le h \sum_{g \in G^{(\le K)}} \lambda(T_g) \sup_{c \in T_g} \sqrt{A(c) \det \Sigma_\theta(c)} \\ &= h \sum_{g \in G^{(\le K)}} \mathrm{Vol}_g\, \widehat{\mathcal{J}}_g \le h \left|G^{(\le K)}\right| \widehat{V}\, \widehat{\mathcal{J}}. \end{aligned} \tag{80}$$

By (80), the Riemannian volume, and every covering-number prefactor built from it, grows at most linearly in the cardinality $|G^{(\le K)}|$ of the Poisson DAG. □

Corollary A.9 (DAG–sized metric entropy of the elementary-flow set)**.** *For every radius $0 < \delta \le \frac{1}{2}\mathrm{Inj}(\mathscr{S}_h^\Theta)$, the δ-covering number of $\mathscr{S}_h^\Theta$ in the left-invariant geodesic metric d_G satisfies*

$$N\big(\delta, \mathscr{S}_h^\Theta, d_G\big) \le \frac{2^{d_\mathscr{S}} h \mathrm{Vol}_g \widehat{\mathcal{J}}}{c_\mathscr{S}} \big|G^{(\le K)}\big| \delta^{-d_\mathscr{S}}, \quad d_\mathscr{S} = d + 1,$$

where Vol_g is the common Lebesgue volume of each tile T_g, and $\widehat{\mathcal{J}} = \sqrt{A_0 \det \Sigma_0}$ as defined above.

Proof. For tiles of common Lebesgue volume, $\widehat{V} = \mathrm{Vol}_g$, and Proposition A.8 with $d_\mathscr{S} = d + 1$ coordinates (ε, c) gives

$$\mathrm{Vol}(\mathscr{S}_h^\Theta) \le h \left|G^{(\le K)}\right| \mathrm{Vol}_g\, \widehat{\mathcal{J}}. \tag{81}$$

Inserting (81) into Proposition A.7, for $0 < \delta \le \frac{1}{2}\mathrm{Inj}(\mathscr{S}_h^\Theta)$,

$$N\big(\delta, \mathscr{S}_h^\Theta, d_G\big) \le \frac{2^{d_\mathscr{S}} \mathrm{Vol}(\mathscr{S}_h^\Theta)}{c_\mathscr{S}}\, \delta^{-d_\mathscr{S}} \le \frac{2^{d_\mathscr{S}} h \mathrm{Vol}_g\, \widehat{\mathcal{J}}}{c_\mathscr{S}}\, \left|G^{(\le K)}\right|\, \delta^{-d_\mathscr{S}}, \tag{82}$$

which is the claim. Combining (82) with Proposition A.5 under the uniform allocation $\delta_j = \varepsilon/M$, valid whenever $\varepsilon/M \le \frac{1}{2}\mathrm{Inj}(\mathscr{S}_h^\Theta)$, yields the walk-level entropy

$$N\big(\varepsilon, \mathcal{W}_{r,K,M,h}^\theta, d_G\big) \le \Big[N\big(\tfrac{\varepsilon}{M}, \mathscr{S}_h^\Theta, d_G\big)\Big]^M \le \left[\frac{2^{d_\mathscr{S}} h \mathrm{Vol}_g\, \widehat{\mathcal{J}}}{c_\mathscr{S}} \left|G^{(\le K)}\right| \Big(\frac{M}{\varepsilon}\Big)^{d_\mathscr{S}}\right]^M,$$

which matches the informal scaling discussed above. □

Proposition A.10 (Uniform trajectory-risk deviation)**.** *Let $\mathcal{F}_K$ be the depth-K Neural–Poisson Lie-walk class with $M = \lceil T/\epsilon \rceil$ steps on a compact domain $\mathcal{M} \subset \mathbb{R}^{2n}$ of Euclidean diameter D_2, and let L and $\widehat{L}_n$ be the population and empirical squared trajectory risks. Then, for every $\delta \in (0,1)$, with probability at least $1 - \delta$ over an i.i.d. sample of size n,*

$$\sup_{W \in \mathcal{F}_K} \big|L(W) - \widehat{L}_n(W)\big| \le \frac{32 D_2^2}{\sqrt{n}} \sqrt{M\Big[\log\big(C_1 |\mathcal{G}^{(\le K)}|\big) + d_\mathscr{S} \log\Big(\frac{2 D_2 M n}{D_2^2}\Big)\Big]} + D_2^2 \sqrt{\frac{2\log(2/\delta)}{n}},$$

with the covering constant C_1 of Corollary A.9.

Proof of Proposition 4.2. Let $\mathcal{M} \subset \mathbb{R}^{2n}$ be compact with Euclidean diameter $D_2 := \sup_{x,y\in\mathcal{M}} \|x-y\|_2$. Fix integers $K \geq 0$ and $M = \lceil T/\epsilon \rceil$ for some $\epsilon > 0$, and let $\mathcal{F}_K = \mathcal{W}^{\theta}_{r(K),K,M,h_{\max}} \subset \mathscr{T}_{r(K),K}$. For $W \in \mathcal{F}_K$ and an i.i.d. sample $\{(Z_{0,i}, Z_i)\}_{i=1}^n$, the squared Euclidean trajectory loss and the population and empirical risks are

$$\ell(W; z_0, z) := \|W(z_0) - z\|_2^2, \quad L(W) := \mathrm{E}[\ell(W; Z_0, Z)], \quad \widehat{L}_n(W) := \frac{1}{n}\sum_{i=1}^{n} \ell\big(W; Z_{0,i}, Z_i\big). \tag{83}$$

The constant $C_1 := 2^{d_{\mathscr{S}}} h \operatorname{Vol}_g \widehat{\mathcal{J}}/c_{\mathscr{S}} > 0$ is the universal covering constant of Corollary A.9 for the elementary flow manifold $\mathscr{S}_{h_{\max}}$, enlarged below where indicated. Throughout, the trajectories remain in $\mathcal{M}$, and $d_\infty^{(2)}(W, W') := \sup_{x\in\mathcal{M}} \|W(x) - W'(x)\|_2$.

For $z_0, z \in \mathcal{M}$ and $W, W' \in \mathcal{F}_K$, the polarization identity $\|u\|_2^2 - \|v\|_2^2 = \langle u - v,\, u + v\rangle$ with $u = W(z_0) - z$, $v = W'(z_0) - z$, the Cauchy–Schwarz inequality, and the triangle inequality give

$$0 \leq \ell(W; z_0, z) = \|W(z_0) - z\|_2^2 \leq D_2^2, \tag{84}$$

$$\begin{aligned} \big|\ell(W; z_0, z) - \ell(W'; z_0, z)\big| &= \big|\big\langle W(z_0) - W'(z_0),\, W(z_0) + W'(z_0) - 2z\big\rangle\big| \\ &\leq \|W(z_0) - W'(z_0)\|_2 \left(\|W(z_0) - z\|_2 + \|W'(z_0) - z\|_2\right) \\ &\leq 2D_2 \,\|W(z_0) - W'(z_0)\|_2 \leq 2D_2\, d_\infty^{(2)}(W, W'). \end{aligned} \tag{85}$$

By (85) the loss is $L = 2D_2$–Lipschitz with respect to $d_\infty^{(2)}$: if $\{W_i\}$ is an (ε/L)-cover of $\mathcal{F}_K$, then $\{\ell(W_i; \cdot, \cdot)\}$ is an ε-cover of $\mathcal{L}_K = \{\ell(W; \cdot, \cdot) : W \in \mathcal{F}_K\}$, and

$$\log N\big(\varepsilon, \mathcal{L}_K, \|\cdot\|_\infty\big) \leq \log N\left(\tfrac{\varepsilon}{L}, \mathcal{F}_K, d_\infty^{(2)}\right). \tag{86}$$

Equip $\mathscr{T}_{r,K}$ with the left–invariant Riemannian structure induced from the L^2 inner product on vector fields in (52). The tangent spaces are finite-dimensional and the walk class is contained in a compact subset of the group; the linear map sending a tangent vector at $\gamma(t)$ to the velocity of the point $\gamma(t)(z)$ is therefore bounded uniformly in t and $z \in \mathcal{M}$, and rescaling the inner product at the identity by a constant normalizes this bound to one (the rescaling multiplies all covering radii by a fixed factor, which is absorbed in C_1). For any smooth curve $\gamma : [0, 1] \to \mathscr{T}_{r,K}$ with $\gamma(0) = W'$, $\gamma(1) = W$ and any $z \in \mathcal{M}$, the normalized speed gives

$$\begin{aligned} \|W(z) - W'(z)\|_2 &= \Big\| \int_0^1 \frac{d}{dt}\big[\gamma(t)(z)\big]\, dt \Big\|_2 \leq \int_0^1 \Big\| \frac{d}{dt}\big[\gamma(t)(z)\big] \Big\|_2 \, dt \\ &\leq \int_0^1 \|\dot\gamma(t)\|_{\gamma(t)}\, dt = L_G(\gamma). \end{aligned} \tag{87}$$

Taking in (87) the infimum over γ and then the supremum over $z \in \mathcal{M}$, and combining with (86),

$$d_\infty^{(2)}(W, W') \leq d_G(W, W') \quad \Longrightarrow \quad \log N\big(\varepsilon, \mathcal{L}_K, \|\cdot\|_\infty\big) \leq \log N\left(\tfrac{\varepsilon}{L}, \mathcal{F}_K, d_G\right). \tag{88}$$

Proposition A.5 with the uniform allocation $\delta_j = \varepsilon/(LM)$, followed by the DAG-sized bound of Proposition A.7 and Corollary A.9 with a constant $C_1 > 0$ absorbing the step ceiling and the geometric constants (radii above $\frac{1}{2}\mathrm{Inj}(\mathscr{S}_h)$ are covered by monotonicity of covering numbers after enlarging C_1, since only radii $\delta \leq D_2/(2M)$ enter below), gives

$$N\left(\frac{\varepsilon}{L}, \mathcal{F}_K, d_G\right) \leq \left[N\left(\frac{\varepsilon}{LM}, \mathscr{S}_h, d_G\right)\right]^M, \tag{89}$$

$$N\left(\delta, \mathscr{S}_h, d_G\right) \leq C_1\, |\mathcal{G}^{(\leq K)}| \Big(\frac{1}{\delta}\Big)^{d_{\mathscr{S}}}, \qquad d_{\mathscr{S}} = d + 1. \tag{90}$$

Substituting $\delta = \varepsilon/(LM)$ in (90), inserting the result in (89), and taking logarithms in (88),

$$\begin{aligned} \log N\big(\varepsilon, \mathcal{L}_K, \|\cdot\|_\infty\big) &\leq M \log N\left(\frac{\varepsilon}{LM}, \mathscr{S}_h, d_G\right) \leq M \log\Big(C_1\, |\mathcal{G}^{(\leq K)}| \Big(\frac{LM}{\varepsilon}\Big)^{d_{\mathscr{S}}}\Big) \\ &= M\left[\log\big(C_1 |\mathcal{G}^{(\leq K)}|\big) + d_{\mathscr{S}} \log\Big(\frac{LM}{\varepsilon}\Big)\right], \qquad L = 2D_2. \end{aligned} \tag{91}$$

By (84) the losses take values in $[0, D_2^2]$, and the truncated Dudley inequality bounds the Rademacher complexity of $\mathcal{L}_K$ by

$$\mathfrak{R}_n(\mathcal{L}_K) \le \inf_{\alpha>0} \left\{ 4\alpha + \frac{12}{\sqrt{n}} \int_{\alpha}^{D_2^2} \sqrt{\log N\big(\varepsilon, \mathcal{L}_K, \|\cdot\|_\infty\big)}\, d\varepsilon \right\}. \tag{92}$$

The square root of the right-hand side of (91) is decreasing,

$$\phi(\varepsilon) := \sqrt{M\left[\log\big(C_1|\mathcal{G}^{(\le K)}|\big) + d_{\mathscr{F}} \log\Big(\frac{LM}{\varepsilon}\Big)\right]}, \qquad \phi'(\varepsilon) = -\frac{M d_{\mathscr{F}}}{2\varepsilon\,\phi(\varepsilon)} < 0, \tag{93}$$

hence, by (91) and (93), for $\alpha \in (0, D_2^2)$

$$\int_{\alpha}^{D_2^2} \sqrt{\log N(\varepsilon, \mathcal{L}_K, \|\cdot\|_\infty)}\, d\varepsilon \le \int_{\alpha}^{D_2^2} \phi(\varepsilon)\, d\varepsilon \le \int_{\alpha}^{D_2^2} \phi(\alpha)\, d\varepsilon = (D_2^2 - \alpha)\,\phi(\alpha). \tag{94}$$

With $A := \log(C_1|\mathcal{G}^{(\le K)}|)$ and $S(\alpha) := \phi(\alpha) = \sqrt{M[A + d_{\mathscr{F}} \log(LM/\alpha)]}$, substituting (94) into (92) gives the one–variable problem

$$\mathfrak{R}_n(\mathcal{L}_K) \le \inf_{\alpha \in (0, D_2^2)} f(\alpha), \qquad f(\alpha) := 4\alpha + \frac{12}{\sqrt{n}} (D_2^2 - \alpha)\, S(\alpha). \tag{95}$$

Differentiating (95) with $S'(\alpha) = -\frac{M d_{\mathscr{F}}}{2\alpha S(\alpha)}$ from (93),

$$f'(\alpha) = 4 + \frac{12}{\sqrt{n}}\Big[-S(\alpha) + (D_2^2 - \alpha) S'(\alpha)\Big] = 4 - \frac{12}{\sqrt{n}}\Big[S(\alpha) + \frac{M d_{\mathscr{F}}}{2S(\alpha)} \cdot \frac{D_2^2 - \alpha}{\alpha}\Big],$$

hence an interior minimizer $\alpha^\star$ satisfies, after multiplication by $\frac{\sqrt{n}}{12}$,

$$S(\alpha^\star) + \frac{M d_{\mathscr{F}}}{2S(\alpha^\star)} \cdot \frac{D_2^2 - \alpha^\star}{\alpha^\star} = \frac{\sqrt{n}}{3}. \tag{96}$$

Instead of solving (96), we bound the infimum in (95) by the value at $\widehat{\alpha} := D_2^2/n$. Enlarging C_1 so that $A \ge 1$, and using $LM/\widehat{\alpha} = 2Mn/D_2 \ge 1$ for $n \ge D_2/(2M)$, gives $S(\widehat{\alpha}) \ge \sqrt{MA} \ge 1$; with $\frac{1}{n} \le \frac{1}{\sqrt{n}}$,

$$\begin{aligned} \mathfrak{R}_n(\mathcal{L}_K) \le f(\widehat{\alpha}) &= \frac{4D_2^2}{n} + \frac{12}{\sqrt{n}}\Big(D_2^2 - \frac{D_2^2}{n}\Big) S(\widehat{\alpha}) \le \frac{4D_2^2}{\sqrt{n}}\, S(\widehat{\alpha}) + \frac{12 D_2^2}{\sqrt{n}}\, S(\widehat{\alpha}) \\ &= \frac{16 D_2^2}{\sqrt{n}} \sqrt{M\left[\log(C_1|\mathcal{G}^{(\le K)}|) + d_{\mathscr{F}} \log\Big(\frac{LMn}{D_2^2}\Big)\right]}. \end{aligned} \tag{97}$$

For losses with values in $[0, D_2^2]$, symmetrization and McDiarmid's inequality [Mohri et al., 2018], applied to $\mathcal{L}_K$ and $-\mathcal{L}_K$ with a union bound, give with probability at least $1 - \delta$

$$\sup_{W \in \mathcal{F}_K} \big|L(W) - \widehat{L}_n(W)\big| \le 2\,\mathfrak{R}_n(\mathcal{L}_K) + D_2^2 \sqrt{\frac{\log(2/\delta)}{2n}}. \tag{98}$$

Inserting (97) with $L = 2D_2$ into (98) and using $\frac{1}{2n} \le \frac{2}{n}$,

$$\begin{aligned} \sup_{W \in \mathcal{F}_K} \big|L(W) - \widehat{L}_n(W)\big| \le{}& \frac{32 D_2^2}{\sqrt{n}} \sqrt{M\left[\log(C_1|\mathcal{G}^{(\le K)}|) + d_{\mathscr{F}} \log\Big(\frac{2D_2 Mn}{D_2^2}\Big)\right]} \\ &+ D_2^2 \sqrt{\frac{2\log(2/\delta)}{n}}, \end{aligned}$$

which is the claimed uniform deviation bound in the Euclidean-loss formulation. □

A.4 Universal Approximation

In order to rigorously establish the universal approximation capability of the proposed Neural–Poisson DAG framework, we first present a sequence of auxiliary lemmas that characterize its algebraic and approximation properties. These lemmas proved in the subsequent paragraphs provide the necessary building blocks for the final result. The main universal approximation theorem will then follow as a direct consequence of these preparatory results. For simplicity of exposition, we shall assume throughout this section that the underlying manifold is Euclidean, *i.e.*, $\mathcal{M} = \mathbb{R}^{2n}$. Nevertheless, the resulting approximation guarantees extend naturally, with only minor technical modifications, to the setting of general symplectic manifolds.

Lemma A.11 (Every DAG node is an r-th Taylor polynomial)**.** *Fix a truncation order $r \geq 1$ and let $G^{(0)} \subset \mathcal{P}_{\leq r} = \{p : \mathbb{R}^d \to \mathbb{R} \mid \deg p \leq r\}$ denote the initial generators, where $\mathcal{P}_{\leq r}$ is the space of polynomials of total degree $\leq r$. Define the depth–K Neural–Poisson DAG. Then every node $g \in G^{(\leq K)}$ can be written as the r-th Taylor polynomial of some C^s function H around a base point x_0:*

$$g(x) = \pi_{\leq r}\left[\sum_{|\alpha|\leq r} \frac{\partial^\alpha H(x_0)}{\alpha!}(x - x_0)^\alpha\right].$$

Proof. We first show $G^{(\leq K)} \subset \mathcal{P}_{\leq r}$ by induction on the depth $k \geq 0$. For $k = 0$, $G^{(0)} \subset \mathcal{P}_{\leq r}$ by construction. Assume $G^{(k-1)} \subset \mathcal{P}_{\leq r}$ and let $f, g \in G^{(k-1)}$. Multiplication adds degrees, and every term of the Poisson bracket carries one derivative on each factor:

$$\begin{aligned}\deg(fg) &= \deg f + \deg g \leq 2r,\\ \deg\{f,g\} &= \deg\Big(\sum_{i=1}^{n}\big(\partial_{q_i} f\,\partial_{p_i} g - \partial_{p_i} f\,\partial_{q_i} g\big)\Big) \leq (\deg f - 1) + (\deg g - 1) \leq 2r - 2. \end{aligned} \tag{99}$$

By (99), $fg \in \mathcal{P}_{\leq 2r}$ and $\{f, g\} = \nabla f^\top J \nabla g \in \mathcal{P}_{\leq 2r-2}$, and the truncation maps both back to degree at most r,

$$\deg \pi_{\leq r}(fg) \leq r, \qquad \deg \pi_{\leq r}\{f, g\} \leq r, \tag{100}$$

hence every node created at depth k lies in $\mathcal{P}_{\leq r}$, and $G^{(\leq K)} \subset \mathcal{P}_{\leq r}$. Fix $g \in G^{(\leq K)}$, a base point x_0, and set $H := g \in C^\infty$. By (100), $\deg g \leq r$, every derivative of order $r + 1$ vanishes, and Taylor's formula of order r with integral remainder has no remainder:

$$\begin{aligned} g(x) &= \sum_{|\alpha|\leq r} \frac{\partial^\alpha g(x_0)}{\alpha!}(x - x_0)^\alpha + \sum_{|\alpha|=r+1} \frac{r+1}{\alpha!}(x - x_0)^\alpha \int_0^1 (1-t)^r\, \partial^\alpha g\big(x_0 + t(x - x_0)\big)\, dt \\ &= \sum_{|\alpha|\leq r} \frac{\partial^\alpha H(x_0)}{\alpha!}(x - x_0)^\alpha = \pi_{\leq r}\left[\sum_{|\alpha|\leq r} \frac{\partial^\alpha H(x_0)}{\alpha!}(x - x_0)^\alpha\right], \end{aligned} \tag{101}$$

where the last equality holds because the bracket already has degree at most r. The monomials $\{(x - x_0)^\alpha\}_{|\alpha|\leq r}$ form a basis of $\mathcal{P}_{\leq r}$, hence the representation (101) is unique in $\mathcal{P}_{\leq r}$, and the lemma follows. □

Lemma A.12 (Local control of the Taylor remainder in H^s)**.** *Fix a cube Q_ℓ of side h centered at x_ℓ. Let $H \in C^s(Q_\ell)$ and let g_ℓ be the $(s-1)$-st order Taylor polynomial of H at x_ℓ. Then*

$$\|H - g_\ell\|_{H^s(Q_\ell)} \leq C(d, s)\|H\|_{C^s(Q_\ell)}.$$

Proof. Throughout $h \leq 1$, $|x - x_\ell|_\infty \leq h/2$ for $x \in Q_\ell$, and $|Q_\ell| = h^d$. For $|\alpha| \leq s - 1$, differentiation commutes with Taylor expansion: $D^\alpha g_\ell$ is the Taylor polynomial of order $s - 1 - |\alpha|$ of $D^\alpha H$ at x_ℓ. Taylor's formula with integral remainder, the bound $|(x - x_\ell)^\beta| \leq (h/2)^{|\beta|}$,

$\int_0^1 (1-t)^{k-1}dt = \frac{1}{k}$ and the multinomial identity $\sum_{|\beta|=k} \frac{1}{\beta!} = \frac{d^k}{k!}$ give, for $x \in Q_\ell$,

$$\begin{aligned}
\big|D^\alpha (H-g_\ell)(x)\big| &= \Big| \sum_{|\beta|=s-|\alpha|} \frac{s-|\alpha|}{\beta!}(x-x_\ell)^\beta \int_0^1 (1-t)^{s-|\alpha|-1} D^{\alpha+\beta} H\big(x_\ell + t(x-x_\ell)\big)\, dt \Big| \\
&\le \sum_{|\beta|=s-|\alpha|} \frac{s-|\alpha|}{\beta!} \Big(\frac{h}{2}\Big)^{s-|\alpha|} \frac{1}{s-|\alpha|} \max_{|\gamma|=s} \|D^\gamma H\|_{L^\infty(Q_\ell)} \\
&= \frac{d^{\,s-|\alpha|}}{(s-|\alpha|)!} \Big(\frac{h}{2}\Big)^{s-|\alpha|} \max_{|\gamma|=s} \|D^\gamma H\|_{L^\infty(Q_\ell)} \le C(d,s)\, h^{s-|\alpha|}\, \|H\|_{C^s(Q_\ell)}.
\end{aligned} \tag{102}$$

Integrating the square of (102) over Q_ℓ gives (103) for $|\alpha| \le s-1$. For $|\alpha| = s$, the polynomial g_ℓ has degree $s-1$, hence $D^\alpha g_\ell = 0$, which gives (104):

$$\|D^\alpha (H-g_\ell)\|_{L^2(Q_\ell)} \le |Q_\ell|^{1/2}\, \|D^\alpha (H-g_\ell)\|_{L^\infty(Q_\ell)} \le C(d,s)\, h^{s-|\alpha|+\frac{d}{2}}\, \|H\|_{C^s(Q_\ell)}, \tag{103}$$

$$\|D^\alpha (H-g_\ell)\|_{L^2(Q_\ell)} = \|D^\alpha H\|_{L^2(Q_\ell)} \le |Q_\ell|^{1/2}\, \|D^\alpha H\|_{L^\infty(Q_\ell)} \le h^{\frac{d}{2}}\, \|H\|_{C^s(Q_\ell)}. \tag{104}$$

Summing the squares of (103) and (104) over the finitely many multi-indices $|\alpha| \le s$ and using $h^{2(s-|\alpha|)} \le 1$,

$$\begin{aligned}
\|H-g_\ell\|^2_{H^s(Q_\ell)} &= \sum_{|\alpha|\le s} \|D^\alpha (H-g_\ell)\|^2_{L^2(Q_\ell)} \\
&\le \sum_{|\alpha|\le s-1} C(d,s)^2\, h^{2(s-|\alpha|)+d}\, \|H\|^2_{C^s(Q_\ell)} + \sum_{|\alpha|=s} h^d\, \|H\|^2_{C^s(Q_\ell)} \\
&\le C(d,s)^2\, h^d\, \|H\|^2_{C^s(Q_\ell)},
\end{aligned} \tag{105}$$

where the constant is enlarged in the last step and depends only on (d,s), not on the cube index ℓ. Taking square roots in (105) and using $h^{d/2} \le 1$,

$$\|H-g_\ell\|_{H^s(Q_\ell)} \le C(d,s)\, h^{\frac{d}{2}}\, \|H\|_{C^s(Q_\ell)} \le C(d,s)\, \|H\|_{C^s(Q_\ell)}.$$

□

Proposition A.13. *Let $d \ge 1$ and $s > d$ be integers, and set $\Omega = [-1,1]^d \subset \mathbb{R}^d$. For every $H \in C^s(\Omega)$ define the C^s–seminorm $\|H\|_{C^s} := \sum_{|\alpha|\le s} \|\partial^\alpha H\|_{L^\infty(\Omega)}$. Then for each integer $r \ge 1$ there exists a polynomial $\mathrm{Poly}_r \in \mathcal{P}_{\le r}(\mathbb{R}^d)$ and a constant $C(d,s) > 0$ depending only on d and s such that*

$$\|\nabla H - \nabla \mathrm{Poly}_r\|_{L^\infty(\Omega)} \le C(d,s) \|H\|_{C^s} r^{1-s}. \tag{106}$$

Proof. Fix r and set $h := 1/r \le 1$. Cover Ω by the cubes $\{Q_\ell\}$ of side h centred at the points x_ℓ of the grid $\frac{h}{2}\mathbb{Z}^d \cap \Omega$; their interiors cover Ω, and every point of Ω lies in at most 2^d of them. Let $\{\varphi_\ell\}$ be a C^∞ partition of unity subordinate to this cover,

$$\sum_\ell \varphi_\ell \equiv 1 \text{ on } \Omega, \qquad \|\partial^\beta \varphi_\ell\|_{L^\infty} \le C h^{-|\beta|}. \tag{107}$$

For each ℓ, let g_ℓ be the $(s-1)$-st Taylor polynomial of H at x_ℓ (as in Lemma A.11), so that $\partial^\alpha (H-g_\ell)(x_\ell) = 0$ for all $|\alpha| \le s-1$. The pointwise bound (102) for $|\gamma| \le s-1$, and $\partial^\gamma g_\ell = 0$ for $|\gamma| = s$, give for every $x \in Q_\ell$

$$\big|\partial^\gamma (H-g_\ell)(x)\big| \le C(d,s)\, h^{s-|\gamma|}\, \|H\|_{C^s}, \qquad |\gamma| \le s, \tag{108}$$

in particular $|H(x) - g_\ell(x)| \le C h^s \|H\|_{C^s}$ and $|\nabla H(x) - \nabla g_\ell(x)| \le C h^{s-1} \|H\|_{C^s}$. Define the *patched* approximation $\widetilde{P}_r(x) := \sum_\ell \varphi_\ell(x) g_\ell(x)$. By (107), the Leibniz rule, the overlap bound,

and (108), for $x \in \Omega$,

$$\begin{aligned}\big|\nabla H(x) - \nabla \widetilde{P}_r(x)\big| &= \Big|\nabla\Big(\sum_{\ell} \varphi_\ell(x)\big(H(x) - g_\ell(x)\big)\Big)\Big| \\ &\le \sum_{\ell:\, x\in Q_\ell} |\varphi_\ell(x)|\,\big|\nabla H(x) - \nabla g_\ell(x)\big| + \sum_{\ell:\, x\in Q_\ell} |\nabla\varphi_\ell(x)|\,\big|H(x) - g_\ell(x)\big| \\ &\le 2^d\Big(Ch^{s-1}\|H\|_{C^s} + Ch^{-1}\cdot Ch^s\|H\|_{C^s}\Big) \le C_3(d,s)\,h^{s-1}\,\|H\|_{C^s},\end{aligned}$$

and taking the supremum over Ω,

$$\|\nabla H - \nabla \widetilde{P}_r\|_{L^\infty(\Omega)} \le C_3(d,s)\,h^{s-1}\,\|H\|_{C^s}. \tag{109}$$

On Ω, $\widetilde{P}_r = H + \sum_\ell \varphi_\ell\,(g_\ell - H)$. For $|\gamma| \le s$, the Leibniz rule, (107) and (108) give

$$\begin{aligned}\big|\partial^\gamma \widetilde{P}_r(x)\big| &\le \big|\partial^\gamma H(x)\big| + \sum_{\ell:\, x\in Q_\ell}\ \sum_{\beta\le\gamma} \binom{\gamma}{\beta}\big|\partial^\beta\varphi_\ell(x)\big|\,\big|\partial^{\gamma-\beta}(g_\ell - H)(x)\big| \\ &\le \|H\|_{C^s} + 2^d \sum_{\beta\le\gamma}\binom{\gamma}{\beta} Ch^{-|\beta|}\cdot Ch^{s-|\gamma|+|\beta|}\,\|H\|_{C^s} \\ &= \Big(1 + C(d,s)\,h^{s-|\gamma|}\Big)\|H\|_{C^s} \le C(d,s)\,\|H\|_{C^s},\end{aligned}$$

and summing over $|\gamma| \le s$,

$$\|\widetilde{P}_r\|_{C^s} \le C_5(d,s)\,\|H\|_{C^s}. \tag{110}$$

Let Π_r be a standard Jackson/de la Vallée–Poussin projector onto $\mathcal{P}_{\le r}$ (total degree $\le r$), which satisfies the derivative Jackson estimate

$$\|\nabla(f - \Pi_r f)\|_{L^\infty(\Omega)} \le C_4(d,s)\,r^{1-s}\,\|f\|_{C^s(\Omega)} \qquad \text{for all } f \in C^s(\Omega). \tag{111}$$

For the polynomial $\mathrm{Poly}_r := \Pi_r(\widetilde{P}_r) \in \mathcal{P}_{\le r}$, the triangle inequality, (109), (111), (110) and $h = 1/r$ give

$$\begin{aligned}\|\nabla H - \nabla \mathrm{Poly}_r\|_{L^\infty(\Omega)} &\le \|\nabla H - \nabla\widetilde{P}_r\|_{L^\infty(\Omega)} + \|\nabla(\widetilde{P}_r - \Pi_r\widetilde{P}_r)\|_{L^\infty(\Omega)} \\ &\le C_3\,h^{s-1}\|H\|_{C^s} + C_4\,r^{1-s}\|\widetilde{P}_r\|_{C^s} \\ &\le C_3\,r^{1-s}\|H\|_{C^s} + C_4C_5\,r^{1-s}\|H\|_{C^s} = C(d,s)\,r^{1-s}\,\|H\|_{C^s},\end{aligned}$$

which is exactly (106). □

Lemma A.14 (Supremum-Norm Bound via Poisson–Bracket Coefficients)**.** *Let $M \subset \mathbb{R}^{2n}$ be a compact phase–space domain equipped with canonical coordinates $(q_1,\dots,q_n,p_1,\dots,p_n)$. For a Hamiltonian $g \in C^{r+1}(M)$ with $r \ge 1$, denote the associated Hamiltonian vector field by*

$$X_g = \sum_{i=1}^{n}\big(\partial_{p_i} g\,\partial_{q_i} - \partial_{q_i} g\,\partial_{p_i}\big), \qquad \mathrm{ad}_g^k(\cdot) := \underbrace{\{\dots\{\{\cdot, g\}, g\}, \dots, g\}}_{k \text{ times}}.$$

Given a fixed step size $\epsilon > 0$, define the local Lie–series remainder by

$$R_\epsilon^{r+1}(z) = \frac{\epsilon^{r+1}}{r!}\int_0^1 (1-\tau)^r X_g^{r+1}\big(e^{\tau\epsilon X_g}(z)\big)\,d\tau, \qquad z \in M.$$

Then there exists a constant $C(n,r) > 0$ depending only on n and r such that

$$\big\|R_\epsilon^{r+1}\big\|_{L^\infty(M)} \le \frac{\epsilon^{r+1}}{(r+1)!} C(n,r)\|g\|_{C^{r+1}(M)}^{r+1}.$$

Remark. Up to this point some auxiliary bounds were stated on $\Omega = [-1,1]^d$. The present estimate is domain-agnostic: the coefficient counting for $X_g^k z^\beta = \mathrm{ad}_g^k(z^\beta)$ only produces a finite sum of

products of k partial derivatives from $\{\partial^\alpha g : 1 \le |\alpha| \le k\}$ with universal combinatorial coefficients. Hence replacing $\|\cdot\|_{L^\infty(\Omega)}$ by $\|\cdot\|_{L^\infty(M)}$ leaves the bound unchanged. If one prefers a reduction: cover the compact M by finitely many axis-aligned cubes Q_ℓ affine to Ω; since our inequality uses only suprema of derivatives of g (no Jacobian weights), taking $\max_\ell \|\partial^\alpha g\|_{L^\infty(Q_\ell)} = \|\partial^\alpha g\|_{L^\infty(M)}$ yields the same constants.

Proof. Let z^β, $\beta = 1, \dots, 2n$, denote a canonical coordinate. We show by induction on $k \ge 1$ that $X_g^k z^\beta = \mathrm{ad}_g^k(z^\beta)$ is a sum of at most $(2n)^{k-1}(k-1)!$ terms of the form

$$\pm \prod_{i=1}^{k} \partial_z^{\alpha_i} g, \qquad 1 \le |\alpha_i| \le k. \tag{112}$$

For $k = 1$, $X_g q_i = \partial_{p_i} g$ and $X_g p_i = -\partial_{q_i} g$ consist of a single factor. Assume (112) for k. Since $X_g = \sum_l \big(\partial_{p_l} g\, \partial_{q_l} - \partial_{q_l} g\, \partial_{p_l}\big)$ is a derivation, the Leibniz rule gives, for one term of $X_g^k z^\beta$,

$$\begin{aligned} X_g\Big(\prod_{i=1}^{k} \partial_z^{\alpha_i} g\Big) &= \sum_{i=1}^{k} \Big(X_g\, \partial_z^{\alpha_i} g\Big) \prod_{j \ne i} \partial_z^{\alpha_j} g \\ &= \sum_{i=1}^{k} \sum_{l=1}^{n} \Big(\partial_{p_l} g\, \partial_{q_l} \partial_z^{\alpha_i} g - \partial_{q_l} g\, \partial_{p_l} \partial_z^{\alpha_i} g\Big) \prod_{j \ne i} \partial_z^{\alpha_j} g. \end{aligned} \tag{113}$$

Each of the $2nk$ terms in (113) is a product of $k+1$ derivatives of g of orders between 1 and $k+1$, hence $X_g^{k+1} z^\beta$ has at most $2nk \cdot (2n)^{k-1}(k-1)! = (2n)^k k!$ terms of the form (112). Bounding every factor by $\|\partial_z^\alpha g\|_{L^\infty(M)} \le \|g\|_{C^k(M)}$,

$$\big\|X_g^k z^\beta\big\|_{L^\infty(M)} \le (2n)^{k-1}(k-1)!\, \|g\|_{C^k(M)}^{k}, \qquad \big\|X_g^{r+1} z^\beta\big\|_{L^\infty(M)} \le (2n)^r\, r!\, \|g\|_{C^{r+1}(M)}^{r+1}. \tag{114}$$

Along the Hamiltonian flow $\phi_t = e^{tX_g}$, the function $t \mapsto z^\beta(\phi_t(z))$ has derivatives $\frac{d^k}{dt^k} z^\beta(\phi_t(z)) = (X_g^k z^\beta)(\phi_t(z))$. Taylor's formula in t of order r with integral remainder, and the substitution $t = \tau\epsilon$, give

$$\begin{aligned} z^\beta\big(\phi_\epsilon(z)\big) &= \sum_{k=0}^{r} \frac{\epsilon^k}{k!} \big(X_g^k z^\beta\big)(z) + \frac{1}{r!} \int_0^\epsilon (\epsilon - t)^r \, \big(X_g^{r+1} z^\beta\big)\big(\phi_t(z)\big)\, dt \\ &= \sum_{k=0}^{r} \frac{\epsilon^k}{k!} \big(X_g^k z^\beta\big)(z) + \underbrace{\frac{\epsilon^{r+1}}{r!} \int_0^1 (1-\tau)^r \, \big(X_g^{r+1} z^\beta\big)\big(\phi_{\tau\epsilon}(z)\big)\, d\tau}_{=: R_{\epsilon,\beta}^{r+1}(z)}. \end{aligned} \tag{115}$$

For $z \in M$ the flow remains in M for times in $[0, \epsilon]$, hence the supremum over M can be taken inside the integrand of (115). With $\int_0^1 (1-\tau)^r d\tau = \frac{1}{r+1}$ and (114),

$$\|R_{\epsilon,\beta}^{r+1}\|_{L^\infty(M)} \le \frac{\epsilon^{r+1}}{r!} \int_0^1 (1-\tau)^r\, d\tau\, \big\|X_g^{r+1} z^\beta\big\|_{L^\infty(M)} \le \frac{\epsilon^{r+1}}{(r+1)!}\, (2n)^r\, r!\, \|g\|_{C^{r+1}(M)}^{r+1}. \tag{116}$$

The bound (116) holds uniformly for each coordinate index $\beta = 1, \dots, 2n$; for the Euclidean norm of the vector-valued remainder $R_\epsilon^{r+1} = (R_{\epsilon,1}^{r+1}, \dots, R_{\epsilon,2n}^{r+1})$ the constant acquires the factor $\sqrt{2n}$, and $C(n,r) := \sqrt{2n}\,(2n)^r\, r!$ completes the proof. □

Theorem A.15 (Symplectic Lie Transform Error Expansion)**.** *Let $H \in C^s(\mathbb{R}^{2n})$ with $s > 2n$, let $X_H = J\nabla H$ be the exact Hamiltonian vector field, and let P_r be an r-th degree polynomial approximation of H with associated vector field $X_\epsilon^H := J\nabla P_r$. Define the exact flows*

$$\hat{\Phi}_{\epsilon,H}(z) := \exp(\epsilon X_H)(z), \qquad F_\epsilon^{X_\epsilon^H}(z) := \exp(\epsilon X_\epsilon^H)(z).$$

Assume X_H and X_ϵ^H are (globally) Lipschitz on M with common constant L. Then, for all sufficiently small $\epsilon > 0$,

$$\big\|\hat{\Phi}_{\epsilon,H} - F_\epsilon^{X_\epsilon^H}\big\|_{L^\infty} \le C(s,n,r) \left[r^{1-s} \|H\|_{C^s} \epsilon e^{\epsilon L} + \frac{\epsilon^{r+1}}{(r+1)!} \|H\|_{C^s}^{r+1} \right]. \tag{117}$$

Proof. All norms are taken over a compact domain $M \subset \Omega = [-1,1]^{2n}$ that contains the flow lines for times in $[0,\epsilon]$. We take for P_r the polynomial Poly_r of Proposition A.13 with $d = 2n$, shifted by a constant so that $P_r(x_0) = H(x_0)$ at some $x_0 \in \Omega$; constants leave $X_\epsilon^H = J\nabla P_r$ unchanged. Exact flows are written $e^{tX}(z)$, and, as for the drivers of $\mathscr{T}_{r,K}$, the flow generated by the polynomial field is evaluated through its order-r Lie series, whose expansion carries no remainder:

$$F_\epsilon^{X_\epsilon^H}(z) = z + \sum_{k=1}^{r} \frac{\epsilon^k}{k!} X_{P_r}^k(z). \tag{118}$$

Inserting the exact flow $e^{\epsilon X_{P_r}}$ of the polynomial field into (118),

$$\big\|\hat\Phi_{\epsilon,H}(z) - F_\epsilon^{X_\epsilon^H}(z)\big\| \le \underbrace{\big\|e^{\epsilon X_H}(z) - e^{\epsilon X_{P_r}}(z)\big\|}_{\text{model error}} + \underbrace{\Big\|e^{\epsilon X_{P_r}}(z) - z - \sum_{k=1}^{r} \frac{\epsilon^k}{k!} X_{P_r}^k(z)\Big\|}_{\text{Lie-series remainder}}. \tag{119}$$

Both flows solve their integral equations from the same initial point. Adding and subtracting $X_H\big(e^{\tau X_{P_r}}(z)\big)$ and using the common Lipschitz constant L, for $t \in [0,\epsilon]$,

$$\begin{aligned}
\big\|e^{tX_H}(z) - e^{tX_{P_r}}(z)\big\| &= \Big\|\int_0^t \Big(X_H\big(e^{\tau X_H}(z)\big) - X_{P_r}\big(e^{\tau X_{P_r}}(z)\big)\Big)d\tau\Big\| \\
&\le \int_0^t \big\|X_H\big(e^{\tau X_H}(z)\big) - X_H\big(e^{\tau X_{P_r}}(z)\big)\big\|\, d\tau \\
&\quad + \int_0^t \big\|X_H\big(e^{\tau X_{P_r}}(z)\big) - X_{P_r}\big(e^{\tau X_{P_r}}(z)\big)\big\|\, d\tau \\
&\le L\int_0^t \big\|e^{\tau X_H}(z) - e^{\tau X_{P_r}}(z)\big\|\, d\tau + t\,\|X_H - X_{P_r}\|_{L^\infty}.
\end{aligned} \tag{120}$$

The inhomogeneous term $t\,\|X_H - X_{P_r}\|_{L^\infty}$ in (120) is nondecreasing in t, and Grönwall's inequality, $e^x - 1 \le xe^x$, $J^\top J = I_{2n}$ and (106) bound the model error in (119):

$$\begin{aligned}
\big\|e^{\epsilon X_H}(z) - e^{\epsilon X_{P_r}}(z)\big\| &\le \frac{e^{\epsilon L}-1}{L}\,\|X_H - X_{P_r}\|_{L^\infty} \le \epsilon\, e^{\epsilon L}\,\big\|J(\nabla H - \nabla P_r)\big\|_{L^\infty} \\
&= \epsilon\, e^{\epsilon L}\,\|\nabla H - \nabla P_r\|_{L^\infty} \le C(2n,s)\, r^{1-s}\,\|H\|_{C^s}\,\epsilon\, e^{\epsilon L}.
\end{aligned} \tag{121}$$

The Lie-series remainder in (119) is the integral remainder R_ϵ^{r+1} of the Lie series of P_r, and Lemma A.14 with $g = P_r$ gives

$$\Big\|e^{\epsilon X_{P_r}}(z) - z - \sum_{k=1}^{r} \frac{\epsilon^k}{k!} X_{P_r}^k(z)\Big\| \le \frac{\epsilon^{r+1}}{(r+1)!}\, C(n,r)\,\|P_r\|_{C^{r+1}}^{r+1}. \tag{122}$$

Markov's inequality $\|p'\|_{L^\infty[-1,1]} \le r^2\|p\|_{L^\infty[-1,1]}$ for univariate polynomials of degree $\le r$, applied along coordinate lines of Ω, the mean value theorem with $P_r(x_0) = H(x_0)$ and $\operatorname{diam}\Omega = 2\sqrt{2n}$, and (106) control the norm in (122):

$$\begin{aligned}
\|P_r\|_{C^{r+1}} &= \sum_{|\gamma|\le r+1} \|\partial^\gamma P_r\|_{L^\infty} \le \sum_{|\gamma|\le r+1} r^{2|\gamma|}\,\|P_r\|_{L^\infty} = C(n,r)\,\|P_r\|_{L^\infty} \\
&\le C(n,r)\Big(\|H\|_{L^\infty} + \sup_{x\in\Omega}\big|(H-P_r)(x) - (H-P_r)(x_0)\big|\Big) \\
&\le C(n,r)\Big(\|H\|_{L^\infty} + 2\sqrt{2n}\,\|\nabla H - \nabla P_r\|_{L^\infty}\Big) \\
&\le C(n,r)\Big(1 + 2\sqrt{2n}\,C(2n,s)\,r^{1-s}\Big)\|H\|_{C^s} \le C(s,n,r)\,\|H\|_{C^s}.
\end{aligned} \tag{123}$$

Inserting (121), (122) and (123) into (119) and taking the supremum over $z \in M$,

$$\begin{aligned}
\big\|\hat\Phi_{\epsilon,H} - F_\epsilon^{X_\epsilon^H}\big\|_{L^\infty} &\le C(2n,s)\,r^{1-s}\|H\|_{C^s}\,\epsilon\, e^{\epsilon L} + \frac{\epsilon^{r+1}}{(r+1)!}\,C(n,r)\,C(s,n,r)^{r+1}\,\|H\|_{C^s}^{r+1} \\
&\le C(s,n,r)\left[r^{1-s}\|H\|_{C^s}\,\epsilon\, e^{\epsilon L} + \frac{\epsilon^{r+1}}{(r+1)!}\,\|H\|_{C^s}^{r+1}\right],
\end{aligned}$$

with $C(s,n,r)$ enlarged in the last step; the constant of the first term, $C(2n,s)$, is independent of r. This establishes (117). □

Proposition A.16 (Global error for an M–step Lie–group walk)**.** *Let $H \in C^s(\mathbb{R}^{2n})$ with $s > 2n$, let P_r be a degree–r polynomial approximant of H, and let F_T be the exact time-T flow of $X_H = J\nabla H$. If every step of size ϵ satisfies the one–step bound of Theorem A.15, and X_H and X_{P_r} share a Lipschitz bound L_* on the relevant compact set, then the M–step walk with $T = M\epsilon$ obeys*

$$\|\mathcal{W}_{\mathbf{g}}(M) - F_T\|_{L^\infty} \le \frac{e^{L_* T} - 1}{L_*} \Big[C_1(s,n,r) r^{1-s} \|H\|_{C^s} e^{\epsilon L_*} + \frac{C_2(s,n,r)}{(r+1)!} \|H\|_{C^s}^{r+1} \epsilon^r \Big]. \tag{124}$$

Proof. Let $X_H = J\nabla H$ and $X_{P_r} = J\nabla P_r$, let G_ϵ be the one-step map of the walk, so that $\mathcal{W}_{\mathbf{g}}(M) = G_\epsilon^M$, and let $F_t := \exp(tX_H)$ be the exact flow, with $F_T = \exp(TX_H)$ and $T = M\epsilon$. The one–step bound of Theorem A.15 reads

$$\|G_\epsilon - F_\epsilon\|_{L^\infty} \le \underbrace{C_1(s,n,r) r^{1-s} \|H\|_{C^s}}_{:=\mathfrak{A}} \epsilon e^{\epsilon L_*} + \underbrace{\frac{C_2(s,n,r)}{(r+1)!} \|H\|_{C^s}^{r+1}}_{:=\mathfrak{B}} \epsilon^{r+1} =: \delta_\epsilon. \tag{125}$$

Introduce the accumulated global error after k steps,

$$\mathrm{error}_k := \big\| G_\epsilon^k - F_{k\epsilon} \big\|_{L^\infty}, \qquad k = 0, 1, \dots, M, \qquad \mathrm{error}_0 = \big\| \mathrm{Id} - \mathrm{Id} \big\|_{L^\infty} = 0. \tag{126}$$

The supremum is taken over the compact forward-invariant set on which the one-step bound holds, so that the one-step bound applies at every point $G_\epsilon^k(z)$ visited by the walk. For x, y in this set, the integral equation of the flow and the L_*–Lipschitz continuity of X_H give

$$\begin{aligned} \big\| F_t(x) - F_t(y) \big\| &\le \|x - y\| + \int_0^t \big\| X_H(F_\tau(x)) - X_H(F_\tau(y)) \big\| \, d\tau \\ &\le \|x - y\| + L_* \int_0^t \big\| F_\tau(x) - F_\tau(y) \big\| \, d\tau, \end{aligned} \tag{127}$$

and Grönwall's inequality applied to (127) yields

$$\big\| F_\epsilon(x) - F_\epsilon(y) \big\| \le e^{L_* \epsilon} \|x - y\|, \qquad \|F_\epsilon\|_{\mathrm{Lip}} \le e^{L_* \epsilon}. \tag{128}$$

Adding and subtracting the mixed term $F_\epsilon \circ G_\epsilon^k$, which applies one exact step after k discrete steps, and using the triangle inequality, the one-step bound (125) at $y = G_\epsilon^k(z)$, (128) and (126),

$$\begin{aligned} \mathrm{error}_{k+1} &= \big\| G_\epsilon \circ G_\epsilon^k - F_\epsilon \circ F_{k\epsilon} \big\|_{L^\infty} \\ &\le \big\| G_\epsilon \circ G_\epsilon^k - F_\epsilon \circ G_\epsilon^k \big\|_{L^\infty} + \big\| F_\epsilon \circ G_\epsilon^k - F_\epsilon \circ F_{k\epsilon} \big\|_{L^\infty} \\ &\le \sup_y \big\| G_\epsilon(y) - F_\epsilon(y) \big\| + \|F_\epsilon\|_{\mathrm{Lip}} \sup_z \big\| G_\epsilon^k(z) - F_{k\epsilon}(z) \big\| \\ &\le \delta_\epsilon + e^{L_* \epsilon} \, \mathrm{error}_k, \qquad k = 0, 1, \dots, M-1. \end{aligned} \tag{129}$$

The first term in (129) is the local error of the most recent step, and the second term propagates the error accumulated in the first k steps through one exact step. Unrolling (129) from $\mathrm{error}_0 = 0$,

$$\begin{aligned} &\mathrm{error}_1 \le \delta_\epsilon, \\ &\mathrm{error}_2 \le \delta_\epsilon + e^{L_* \epsilon} \, \mathrm{error}_1 \le \delta_\epsilon \big(1 + e^{L_* \epsilon}\big), \\ &\mathrm{error}_3 \le \delta_\epsilon + e^{L_* \epsilon} \, \mathrm{error}_2 \le \delta_\epsilon \big(1 + e^{L_* \epsilon} + e^{2L_* \epsilon}\big), \\ &\mathrm{error}_M \le \delta_\epsilon \sum_{j=0}^{M-1} e^{j L_* \epsilon} = \delta_\epsilon \, \frac{e^{L_* M \epsilon} - 1}{e^{L_* \epsilon} - 1} \le \delta_\epsilon \, \frac{e^{L_* T} - 1}{L_* \epsilon}, \end{aligned} \tag{130}$$

where the general step $\mathrm{error}_{k+1} \le \delta_\epsilon + e^{L_* \epsilon} \delta_\epsilon \sum_{j=0}^{k-1} e^{j L_* \epsilon} = \delta_\epsilon \sum_{j=0}^{k} e^{j L_* \epsilon}$ is an induction on k, and the last inequality uses $T = M\epsilon$ and $e^x - 1 \ge x$ with $x = L_* \epsilon$; (130) is the discrete analogue of the Duhamel formula. Substituting δ_ϵ from (125) and $\mathcal{W}_{\mathbf{g}}(M) = G_\epsilon^M$ into (130),

$$\big\| \mathcal{W}_{\mathbf{g}}(M) - F_T \big\|_{L^\infty} = \mathrm{error}_M \le \frac{e^{L_* T} - 1}{L_* \, \epsilon} \Big[\mathfrak{A} \, \epsilon \, e^{\epsilon L_*} + \mathfrak{B} \, \epsilon^{r+1} \Big] = \frac{e^{L_* T} - 1}{L_*} \Big[\mathfrak{A} \, e^{\epsilon L_*} + \mathfrak{B} \, \epsilon^r \Big],$$

which, with the definitions of $\mathfrak{A}$ and $\mathfrak{B}$, is precisely (124). □

Proposition A.17 (Relative universal approximation on compact sets)**.** *Fix $T > 0$, a compact set $K \subset \mathbb{R}^{2n}$, and $s > 2n$, and assume that the DAG parameterization realizes every polynomial of total degree $\le r$. Then, for every $\delta > 0$ there exist an integer $r \ge 1$ and a step size ϵ such that the M–step Lie–group walk with $M = \lceil T/\epsilon \rceil$ approximates the exact flow uniformly over the C^s-ball $\mathcal{H}_{R,s}$ of radius R:*

$$\sup_{H \in \mathcal{H}_{R,s}} \sup_{z \in K} \left\| \mathcal{W}_{\mathbf{g}}(M)(z) - F_T^{X_H}(z) \right\| < \delta = \mathcal{O}\left(r^{1-s} + \frac{\epsilon^r}{(r+1)!} \right).$$

Proof of Proposition 4.1. The bounded class in the statement, and the Lipschitz bound $L_* > 0$ on the relevant compact forward–invariant set for the target and DAG polynomial vector fields $X_H = J\nabla H$ and $X_{P_r} = J\nabla P_r$, are

$$\begin{aligned} \mathcal{H}_{R,s} &:= \{H \in C^s(\mathbb{R}^{2n}) : \|H\|_{C^s} \le R\}, \\ \mathrm{Lip}(X_H) \le L_*, \quad \mathrm{Lip}(X_{P_r}) &\le L_*, \qquad H \in \mathcal{H}_{R,s},\ r \in \mathbb{N}, \end{aligned} \tag{131}$$

and step sizes range over $\epsilon \in (0, \epsilon_0]$. Fix $\delta > 0$ and $H \in \mathcal{H}_{R,s}$. For an integer $r \ge 1$, let P_r be the degree–r polynomial of Theorem A.15, realized by the DAG by assumption. Proposition A.16 applied to the pair (H, P_r) on K, with $T = M\epsilon$ and the common Lipschitz constant L_* of (131), and the bound $\|H\|_{C^s} \le R$ give

$$\begin{aligned} \left\| \mathcal{W}_{\mathbf{g}}(M) - F_T^{X_H} \right\|_{L^\infty(K)} &\le \frac{e^{L_* T} - 1}{L_*} \left[C_1(s,n,r)\, r^{1-s} \|H\|_{C^s}\, e^{\epsilon L_*} + \frac{C_2(s,n,r)}{(r+1)!} \|H\|_{C^s}^{r+1} \epsilon^r \right] \\ &\le \frac{e^{L_* T} - 1}{L_*} \left[C_1(s,n,r)\, R\, r^{1-s}\, e^{\epsilon L_*} + \frac{C_2(s,n,r)}{(r+1)!} R^{r+1} \epsilon^r \right]. \end{aligned} \tag{132}$$

By (121), C_1 is the constant of (106) and depends only on (s, n). The two terms of (132) are made small in sequence: first r, then ϵ. Restrict to $\epsilon \le \ln 2 / L_*$, so that $e^{\epsilon L_*} \le 2$. Since $s > 2n \ge 2$, $r^{1-s} \to 0$ as $r \to \infty$, and there exists $r \ge 1$ with

$$\frac{e^{L_* T} - 1}{L_*}\, C_1\, R\, r^{1-s}\, e^{\epsilon L_*} \le \frac{e^{L_* T} - 1}{L_*}\, 2\, C_1\, R\, r^{1-s} \le \frac{\delta}{3}. \tag{133}$$

With this r fixed, $\epsilon^r \to 0$ as $\epsilon \to 0$; choose an integer M large enough that $\epsilon := T/M \le \min\{\epsilon_0, \ln 2 / L_*\}$ and

$$\frac{e^{L_* T} - 1}{L_*}\, \frac{C_2(s,n,r)}{(r+1)!}\, R^{r+1}\, \epsilon^r \le \frac{\delta}{3}, \qquad M = \left\lceil \frac{T}{\epsilon} \right\rceil, \quad T = M\epsilon. \tag{134}$$

Inserting (133) and (134) into (132),

$$\left\| \mathcal{W}_{\mathbf{g}}(M) - F_T^{X_H} \right\|_{L^\infty(K)} \le \frac{\delta}{3} + \frac{\delta}{3} < \delta.$$

The choices (133) and (134) depend only on $(\delta, R, s, n, L_*, T)$ and the fixed constants of Proposition A.16; the same pair (r, ϵ) therefore works uniformly for all $H \in \mathcal{H}_{R,s}$. Taking the supremum over $z \in K$ and $H \in \mathcal{H}_{R,s}$ proves the claim, with the bound of order $r^{1-s} + \epsilon^r/(r+1)!$ in the bracket of (132). □

Algorithm 1 CoefNet$_\theta$: Neural–Polynomial Coefficient Network

Require: Data dimensionality n, Batch size B, horizon T, degree size d_r; normalized time grid $\mathbf{t} \in [0,1]^{B\times T\times 1}$; flattened initial state $x_0 \in \mathbb{R}^{B\times d_s}$.

Ensure: Coefficient tensor $C_\theta \in \mathbb{R}^{B\times T\times d_r\times 2n}$.

1: **Time embedding**: for each (b,t) compute $\tau_{b,t} \leftarrow \phi_t(t_{b,t}) \in \mathbb{R}^{d_{\text{model}}}$, where ϕ_t is an MLP.

2: **State embedding**: for each b compute $s_b \leftarrow \phi_s(x_{0,b}) \in \mathbb{R}^{d_{\text{model}}}$, where ϕ_s is an MLP.

3: **Token sequence**: form $T_b \leftarrow \big[s_b,\ \tau_{b,1},\dots,\tau_{b,T}\big] \in \mathbb{R}^{(T+1)\times d_{\text{model}}}$ for each b; stack to $T \in \mathbb{R}^{(T+1)\times B\times d_{\text{model}}}$.

4: **Transformer encoding**: $E \leftarrow \text{Encoder}_\theta(T) \in \mathbb{R}^{(T+1)\times B\times d_{\text{model}}}$ ▷ stack of K encoder layers with width d_{model}

5: **Discard state token**: $E_{1:T} \leftarrow E[2:] \in \mathbb{R}^{T\times B\times d_{\text{model}}}$.

6: **Projection to coefficients**: for each (t,b) compute

$$a_{t,b} \leftarrow \text{Proj}_\theta\big(E_{t,b}\big) \in \mathbb{R}^{d_r\cdot 6}, \qquad \text{Proj}_\theta = \begin{cases} \text{Linear}(d_{\text{model}} \to d_r\cdot 6), & \text{single layer,} \\ \text{Linear} \to \text{GELU} \to \text{Linear}, & \text{low rank option.} \end{cases}$$

7: **Reshape**: $C_\theta[b,t,:,:] \leftarrow \text{reshape}\big(a_{t,b}, d_r, 2n\big)$ for all (b,t).

8: **return** C_θ.

A.5 Algorithm

In this subsection, we present the algorithmic components that realize the proposed Neural–Poisson framework in a form suitable for direct implementation. Each algorithm corresponds to a distinct stage of the overall pipeline: the construction of the truncated polynomial generator set and its Lie–algebraic structure, the generation of time–conditioned Hamiltonian coefficients by a neural network, the parallel composition of elementary flows via an associative scan, and the high–order symplectic integration of the resulting vector fields. Together, these procedures provide a complete, self–contained specification of the model from algebraic preprocessing to numerical time stepping.

Neural-Polynomial Coefficient Network. Algorithm 1 specifies the coefficient–generation module $\textsc{CoefNet}_\theta$, which defines the parametrization map from the conditioning variables to the coefficients of a truncated polynomial Hamiltonian. Given the normalized time grid and the flattened initial state, the network produces the coefficient tensor $C_\theta \in \mathbb{R}^{B\times T\times d_r\times 6}$ whose entries determine the time–dependent polynomial force field employed in the subsequent symplectic integration.

The procedure begins by embedding each time token $t_{b,t}$ into a latent vector $\tau_{b,t} = \phi_t(t_{b,t}) \in \mathbb{R}^{d_{\text{model}}}$ using a dedicated multi–layer perceptron, and mapping the flattened initial state $x_{0,b}$ into $s_b = \phi_s(x_{0,b}) \in \mathbb{R}^{d_{\text{model}}}$ through a separate multi–layer perceptron. The initial–state embedding is prepended as a special token to the sequence of time embeddings, yielding a tensor in $\mathbb{R}^{(T+1)\times B\times d_{\text{model}}}$. This sequence is processed by a Transformer encoder $\mathrm{Encoder}_\theta$, whose self–attention layers allow each time step to condition on the initial state and to exchange information with other time steps, thereby capturing global temporal dependencies in the coefficient representation. After encoding, the state token is discarded, and each remaining time–step embedding is mapped by a projection head Proj_θ to a vector in $\mathbb{R}^{d_r\cdot 6}$ corresponding to the coefficients of all d_r basis monomials for the six coordinate components. These vectors are then reshaped to form C_θ, which is returned as the network output. This architecture realises a sequence–to–tensor regression map whose inductive biases, consisting of distinct embeddings for time and state, a shared latent representation, and global self–attention, are chosen to respect the structural constraints inherent in Hamiltonian dynamics.

Algorithm 2 BUILDNP_DAG: Neural–Poisson Composition DAG construction

Require: Polynomial degree bound $r \geq 1$, DAG depth $K \geq 0$, local phase dimension d.

Ensure: Nodes $\mathcal{G}^{(\leq K)} = \bigcup_{k=0}^{K} \mathcal{G}^{(k)}$, lookup map $\mathrm{idx} : \mathrm{Exp} \to \{1, \ldots, d_r\}$, and structure tensor $C \in \mathbb{R}^{d_r\times d_r\times d_r}$ satisfying $\{m_i, m_j\} = \sum_m C_{ijm} m_m$ on $P_{\leq r}$.

Construct the truncated monomial basis $P_{\leq r} = \mathrm{span}\{m_\alpha : q^a p^b \mid |a|+|b| \leq r\}$ and its exponent list $\mathrm{Exp} = \{(a, b)\}$ in $\mathbb{N}^{2d}$.
Set $\mathcal{G}^{(0)} \leftarrow P_{\leq r}$ and initialize $\mathcal{G}^{(\leq 0)} \leftarrow \mathcal{G}^{(0)}$.
Build idx by assigning a unique index to each exponent in Exp and set $d_r \leftarrow |\mathrm{Exp}|$.
Initialize $C \leftarrow 0$ in $\mathbb{R}^{d_r\times d_r\times d_r}$.
for $k = 1$ to K **do**
    $\mathcal{C} \leftarrow \emptyset$ ▷ candidate set at layer k
    for each $g \in \mathcal{G}^{(k-1)}$ and each $h \in \mathcal{G}^{(\leq k-1)}$ **do**
        $u \leftarrow \pi_{\leq r}(gh)$ ▷ projected product
        $v \leftarrow \pi_{\leq r}\{g, h\}$ ▷ projected Poisson bracket
        Add u and v to $\mathcal{C}$
    end for
    $\mathcal{G}^{(k)} \leftarrow$ unique elements of $\mathcal{C}$ under exponent equality
    $\mathcal{G}^{(\leq k)} \leftarrow \mathcal{G}^{(\leq k-1)} \cup \mathcal{G}^{(k)}$
end for
Structure tensor C: for monomials $m_i = q^{a_i} p^{b_i}$ and $m_j = q^{a_j} p^{b_j}$,
for $i = 1$ to d_r **do**
    for $j = 1$ to d_r **do**
        for $\ell = 1$ to d **do**
            $s \leftarrow a_i^{(\ell)} b_j^{(\ell)} - a_j^{(\ell)} b_i^{(\ell)}$
            if $s \neq 0$ **then**
                $a \leftarrow a_i + a_j - e_\ell$, $b \leftarrow b_i + b_j - e_\ell$
                if $(a, b) \in \mathrm{Exp}$ with index $m = \mathrm{idx}(a, b)$ **then**
                    $C_{ijm} \leftarrow C_{ijm} + s$
                end if
            end if
        end for
    end for
end for
return $\big(\mathcal{G}^{(\leq K)}, \mathrm{idx}, C\big)$

Neural–Poisson Composition DAG construction. Algorithm 2 describes the construction of the Neural–Poisson composition directed acyclic graph (DAG), which encodes the closure of a truncated polynomial Hamiltonian basis under both pointwise multiplication and the Poisson bracket. The procedure begins by generating the set of all monomials $m_\alpha = q^a p^b$ whose total degree satisfies $|a| + |b| \leq r$, forming the truncated polynomial space $P_{\leq r}$ together with its exponent list Exp. This list is used to initialise the zeroth layer $\mathcal{G}^{(0)}$ of the DAG, which contains all basis monomials, and to build an index map idx assigning a unique integer to each exponent tuple. The integer $d_r = |\mathrm{Exp}|$ records the size of the basis.

Subsequent layers are constructed iteratively up to depth K. At each layer k, all possible projected products $\pi_{\leq r}(gh)$ and projected Poisson brackets $\pi_{\leq r}\{g, h\}$ are computed between generators g from the previous layer $\mathcal{G}^{(k-1)}$ and generators h from all earlier layers $\mathcal{G}^{(\leq k-1)}$. These operations are projected back into $P_{\leq r}$ to preserve the degree bound. The set of distinct resulting monomials, identified by their exponent tuples, forms the new layer $\mathcal{G}^{(k)}$, which is then added to the cumulative set $\mathcal{G}^{(\leq k)}$.

After the generator set is complete, the algorithm computes the structure tensor $C \in \mathbb{R}^{d_r \times d_r \times d_r}$, which encodes the Poisson bracket structure on $P_{\leq r}$. For each ordered pair of basis monomials (m_i, m_j), the integer coefficient $s = a_i^{(\ell)} b_j^{(\ell)} - a_j^{(\ell)} b_i^{(\ell)}$ is evaluated for each coordinate index ℓ, reflecting the antisymmetry of the canonical Poisson structure. When the resulting exponent tuple (a, b) lies in Exp, the corresponding structure constant C_{ijm} is incremented by s. The output of the procedure is the complete generator set $\mathcal{G}^{(\leq K)}$, the exponent–to–index map idx, and the structure tensor C, which together provide an explicit, finite–dimensional Lie algebra representation of the truncated Hamiltonian vector fields.

Algorithm 3 APPLYLIE_R (explicit, even r): r-th order symplectic update via Suzuki–Yoshida composition

---

Require: Current state $z \in \mathbb{R}^{B \times N \times 2d}$, coefficient table $c \in \mathbb{R}^{B \times d_r \times 6}$, exponent tensor Exp, step size $\varepsilon > 0$, target order $r \in 2\mathbb{N}$ (even).

Ensure: Next state $z_+ = \Phi_r(\varepsilon) z$.

1: **Base 2nd–order kernel** *(Störmer–Verlet)*:

$\textsc{KernelSV}(z, h)$:

$g_1 \leftarrow \nabla_{q,p} \hat{H}(z; c, \mathsf{Exp})$

$\dot{q}_1 \leftarrow \partial_p \hat{H}(z), \quad \dot{p}_1 \leftarrow -\partial_q \hat{H}(z), \quad p_{1/2} \leftarrow p + \dfrac{h}{2}\dot{p}_1, \quad q' \leftarrow q + h\dot{q}_1$

$g_2 \leftarrow \nabla_{q,p} \hat{H}\big((q', p_{1/2}); c, \mathsf{Exp}\big), \quad \dot{p}_2 \leftarrow -\partial_q \hat{H}(q', p_{1/2}), \quad p' \leftarrow p_{1/2} + \dfrac{h}{2}\dot{p}_2$

return $z' = (q', p')$

2: **Yoshida/Suzuki coefficient recursion** (order $2 \to 4 \to \cdots \to r$):

- Initialize the coefficient list for order 2: $\mathcal{W}^{(2)} \leftarrow [1]$.
- For $m = 1$ to $r/2 - 1$: set $\alpha \leftarrow 2^{1/(2m+1)}$, $w_1 \leftarrow \dfrac{1}{2 - \alpha}$, $w_0 \leftarrow -\dfrac{\alpha}{2 - \alpha}$, and

$$\mathcal{W}^{(2m+2)} \leftarrow \big[w_1 \mathcal{W}^{(2m)},\ w_0 \mathcal{W}^{(2m)},\ w_1 \mathcal{W}^{(2m)}\big]$$

(concatenation with all entries scaled by the prefactors).

3: **Explicit composition:** set $z^{(0)} \leftarrow z$ and let $\{w_j\}_{j=1}^{S} \leftarrow \mathcal{W}^{(r)}$.

4: **for** $j = 1$ to S **do**

5: $\quad z^{(j)} \leftarrow \textsc{KernelSV}\big(z^{(j-1)}, w_j \varepsilon\big)$

6: **end for**

7: **return** $z_+ \leftarrow z^{(S)}$.

Algorithm 3 implements an explicit rth–order symplectic Lie–integration scheme for truncated polynomial Hamiltonian systems, where r is even. The method is based on the Suzuki–Yoshida

composition of a symmetric second–order kernel, which in this context is given by the Störmer–Verlet splitting of the Hamiltonian vector field.

The input consists of the current state $z = (q, p) \in \mathbb{R}^{B\times N\times 2d}$, the coefficient table c defining the truncated Hamiltonian $\hat{H}(\cdot; c, \mathsf{Exp})$ in the monomial basis indexed by Exp, a time step $\varepsilon > 0$, and the target even order r. The procedure begins by defining the base integrator KERNELSV, which applies one step of the symmetric Störmer–Verlet method to $\hat{H}$ with step size h. This kernel consists of a first half–step update of the momenta (kick), a full–step update of the positions (drift), and a final half–step update of the momenta, where the forces are obtained from the gradient $\nabla_{q,p}\hat{H}$ computed by the truncated polynomial gradient routine.

To reach order $r > 2$, the algorithm constructs the Yoshida–Suzuki composition coefficients $\mathcal{W}^{(r)}$ recursively. Starting from $\mathcal{W}^{(2)} = [1]$ for the base kernel, each stage m is extended to order $2m + 2$ by computing $\alpha = 2^{1/(2m+1)}$ and the weights

$$w_1 = \frac{1}{2-\alpha}, \qquad w_0 = -\frac{\alpha}{2-\alpha},$$

and concatenating the sequence $[w_1\mathcal{W}^{(2m)}, w_0\mathcal{W}^{(2m)}, w_1\mathcal{W}^{(2m)}]$. This recursion preserves the symmetry of the composition and ensures that the local error order increases by two at each step.

In the final composition phase, the method applies the base kernel KERNELSV successively with scaled step sizes $w_j\varepsilon$ for each weight w_j in $\mathcal{W}^{(r)}$, propagating the state from $z^{(0)} = z$ through intermediate states $z^{(j)}$ until $z^{(S)}$, where $S = |\mathcal{W}^{(r)}|$. The output $z_+ = z^{(S)}$ is the state advanced by a single rth–order symplectic step of size ε. This explicit composition preserves the symplectic structure and achieves the prescribed order r for smooth truncated polynomial Hamiltonians, making it suitable for high–accuracy long–time integration in the Neural–Poisson framework.

A.6 Implementation Details

This section specifies the choices required to reproduce our results, including the experimental setup, baseline families, evaluation metrics, runtime and memory protocol, and hyperparameters. It then details the datasets and physical models used in our studies. Unless stated otherwise, all methods share the same splits, physical step size, trajectory length, and device configuration. Any deviations are explicitly highlighted in the corresponding paragraphs and in the appendix.

A.6.1 Experimental Setup

Baselines. We group methods into three families that mirror the table layout in Fig. 2 and Table 1. **(A)** Classical symplectic and high-order integrators using the ground-truth Hamiltonian include Störmer–Verlet [Verlet, 1967], Forest–Ruth 4th order [Forest and Ruth, 1990], and the $s = 2$ Gauss–Legendre collocation scheme [Hairer et al., 2006]. Learned Hamiltonian generators include SymODEN [Zhong et al., 2020], SRNN [Chen et al., 2020], and Nonseparable Symplectic Neural Networks [Xiong et al., 2021]. **(B)** We include KoVAE, a Koopman-regularized variational autoencoder for dynamical data [Naiman et al., 2024], and SKOLR, a structured Koopman-operator linear RNN forecaster [Zhang et al., 2025b]. **(C)** For a line of research in this group, we evaluate EqMotion [Xu et al., 2023], a learned SE(3)-equivariant motion model for atomic trajectories. GeoTDM [Han et al., 2024], a diffusion model over geometric trajectories that enforces SE(3)-equivariance and ET-SEED [Tie et al., 2025], a recent deep generative trajectory model. These methods serve as representative black-box generative baselines that do not incorporate Hamiltonian or Koopman structure.

Evaluation Metrics. We assess prediction accuracy and symplectic structure preservation using four complementary metrics. Prediction accuracy is quantified by the Average and Final Displacement Error, given by $\mathrm{ADE} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T_i}\sum_{t=1}^{T_i}\|z_{i,t} - \hat{z}_{i,t}\|$ and $\mathrm{FDE} = \frac{1}{N}\sum_{i=1}^{N}\|z_{i,T_i} - \hat{z}_{i,T_i}\|$, respectively, where $z_{i,t}$ are predicted states, $\hat{z}_{i,t}$ are ground-truth states, and T_i is the length of the i-th trajectory. Lower values of ADE and FDE indicate superior trajectory prediction accuracy. Structure preservation is evaluated via energy drift (ΔE) and symplectic-form violation ($\Delta\omega$). Energy drift is measured as the average Hamiltonian difference, $\Delta E = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T_i}\sum_{t=1}^{T_i}|H(z_{i,t}) - H(\hat{z}_{i,t})|$. Symplectic-form violation, defined as $\Delta\omega = \|J^\top \nabla J + \nabla J^\top J\|_F$ quantifies deviations from canonical symplectic structure. Lower values of ΔE and $\Delta\omega$ indicate better physical consistency.

Runtime and Memory Measurement. We report *Inference* (seconds per T–step rollout) and, when available, peak GPU memory under a standardized protocol. All experiments were conducted on a single NVIDIA RTX 5090 GPU. Specifically, this value shows the wall–clock time to propagate a single initialization z_0 through a T–step rollout (no backward pass), while the training wall–clock per iteration divided by T (one full rollout + loss + backward + optimizer step, normalized by the physical horizon). All timings are obtained with device–synchronized hardware timers. and the memory is reported as the peak device–resident allocation recorded by the runtime. We average over three runs after a short warm-up, and keep the device type, batch size, trajectory length, and numerical step size identical across methods to ensure fair comparison.

Hyperparameters. Unless otherwise noted, the Poisson basis degree and DAG depth are set to $(r, K) = (5, 2)$, and the model step is matched to the data step, $\varepsilon = \Delta t$. Random seeds are fixed for data generation and training, and validation performance (ADE) is used for early stopping and minor tuning of method-specific regularization where applicable.

A.6.2 Spin Dynamics

In our first experiment, we consider the classical dynamics of spin systems arranged on periodic lattices in one, two, or three dimensions. Each spin is described by a unit vector $\mathbf{S}_i \in \mathbb{S}^2$, representing a classical magnetic moment at lattice site i. A representative atomistic Hamiltonian capturing the principal interactions typically encountered in magnetic simulations and experiments is expressed as follows:

$$H = -\sum_{\langle ij \rangle} J_{ij} \mathbf{S}_i \cdot \mathbf{S}_j + \sum_{\langle ij \rangle} \mathbf{D}_{ij} \cdot (\mathbf{S}_i \times \mathbf{S}_j) - K \sum_i \big(\mathbf{S}_i \cdot \hat{\mathbf{n}}\big)^2 - \sum_i \mathbf{h} \cdot \mathbf{S}_i + \frac{1}{2} \sum_{i \neq j} \mathbf{S}_i^\top \mathbf{T}_{ij} \mathbf{S}_j - \sum_{\langle\langle ij \rangle\rangle} J_{ij}^{(2)} \mathbf{S}_i \cdot \mathbf{S}_j. \tag{135}$$

Each term within this Hamiltonian corresponds to a distinct physical mechanism shaping the static and dynamic properties of spin systems. Below, we briefly describe the nature and physical implications of each interaction.

(i) Heisenberg exchange interaction. The nearest-neighbor Heisenberg exchange interaction, given by the first term $-\sum_{\langle ij \rangle} J_{ij} \mathbf{S}_i \cdot \mathbf{S}_j$, primarily governs the alignment between neighboring spins. For positive exchange coefficients ($J_{ij} > 0$), parallel spin alignment is energetically favored, thus leading to ferromagnetic ordering. Conversely, negative coefficients ($J_{ij} < 0$) promote antiparallel alignment, resulting in antiferromagnetic configurations. This interaction sets the fundamental energy scale for magnetic ordering phenomena and determines short-range correlations and spin-wave excitations observed in spin lattices [Heisenberg, 1928].

(ii) Dzyaloshinskii–Moriya interaction (DMI). The second term, known as the Dzyaloshinskii–Moriya interaction, arises from antisymmetric spin-orbit coupling effects inherent in crystal structures lacking inversion symmetry. Formally represented as $\sum_{\langle ij \rangle} \mathbf{D}_{ij} \cdot (\mathbf{S}_i \times \mathbf{S}_j)$, the DMI promotes non-collinear spin textures exhibiting distinct chirality, such as helices, spirals, and skyrmions. This interaction competes with the symmetric exchange and magnetic anisotropy terms, thus significantly influencing the formation and stabilization of complex spin configurations [Dzyaloshinsky, 1958, Moriya, 1960].

(iii) Magnetocrystalline anisotropy. The magnetocrystalline anisotropy, expressed by the uniaxial term $-K \sum_i (\mathbf{S}_i \cdot \hat{\mathbf{n}})^2$, incorporates crystal-field effects and spin-orbit coupling that energetically distinguish particular crystallographic directions. Depending on the sign of the anisotropy constant K, spins preferentially align either along a particular axis (easy-axis anisotropy for $K > 0$) or within a particular plane (easy-plane anisotropy for $K < 0$). This term strongly influences critical magnetic properties such as coercivity, domain-wall structure, and ferromagnetic resonance frequencies [Aharoni, 1996].

(iv) Zeeman interaction (external field coupling). The Zeeman interaction, represented as $-\sum_i \mathbf{h} \cdot \mathbf{S}_i$, captures the response of magnetic moments to external magnetic fields. This term governs magnetization reversal processes, field-driven hysteresis loops, and spin-wave frequency shifts induced by external magnetic fields. By controlling external fields, experimentalists exploit this coupling to study fundamental spin dynamics and to manipulate spin textures and magnetization states [Evans et al., 2014].

(v) Dipolar (magnetostatic) interaction. The long-range dipolar, or magnetostatic, interaction is encapsulated by the term $\frac{1}{2} \sum_{i \neq j} \mathbf{S}_i^\top \mathbf{T}_{ij} \mathbf{S}_j$, where

$$\mathbf{T}_{ij} = \mu_0 \mu_s^2 r_{ij}^{-3} \big(\mathbf{I} - 3 \hat{\mathbf{r}}_{ij} \hat{\mathbf{r}}_{ij}^\top\big).$$

This interaction accounts for the internal demagnetizing fields arising from spatially distributed magnetic moments. It profoundly influences domain formation, domain-wall structures, and shape-dependent magnetic phenomena such as thin-film magnetism, shape anisotropy, and vortex formation [Aharoni, 1996].

Table 5: **Key simulation settings used for the aspirin and methane molecular dynamics simulations**. Both molecules are parameterized using the OpenFF Sage 2.1.0 force field without cutoffs for nonbonded interactions. Initial atomic coordinates (q_0) are perturbed by Gaussian noise with standard deviations (σ) specified below. Initial atomic momenta (p_0) are sampled from a Maxwell–Boltzmann distribution at temperature $T = 1000$K. Center-of-mass momentum removal is not performed. Trajectories are propagated using the Velocity Verlet integrator with a 1.0fs timestep and run for 1000 integration steps. Atomic coordinates from each integration step are exported in XYZ format for visualization.

	Aspirin	**Methane**
SMILES	`CC(=O)OC1=CC=CC=C1C(=O)O`	`C`
Force field	OpenFF 2.1.0 (main)	OpenFF 2.1.0 (main)
Nonbonded	NoCutoff	NoCutoff
σ (coord. jitter)	0.02	0.01
T for p_0 (K)	1000	1000
COM momentum removal	No	No
Time step (fs)	1.0	1.0
Steps	1000	1000
Trajectory export	XYZ (stride 1)	XYZ (stride 1)

(vi) Next-nearest neighbor exchange (J_2). Finally, the next-nearest neighbor exchange interaction $-\sum_{\langle\langle ij\rangle\rangle} J_{ij}^{(2)} \mathbf{S}_i \cdot \mathbf{S}_j$ introduces competing interactions, thus inducing magnetic frustration in the spin system. The competition between nearest and next-nearest neighbor exchanges (J_1-J_2 models) can lead to complex ground states including spirals, stripes, or even disordered spin liquids, depending sensitively on the relative magnitude of J_2/J_1. Such interactions are critical in understanding frustrated magnetic materials and their associated spin-wave spectra and phase transitions.

A.6.3 Molecular Dynamics

In the experimental results, we study two prototypical molecular systems: **aspirin** (SMILES: `CC(=O)OC1=CC=CC=C1C(=O)O`) and **methane** (SMILES: `C`). For each molecule, we first generate the initial three-dimensional atomic coordinates directly from their respective SMILES strings using RDKit with the ETKDGv3 embedding algorithm. Hydrogen atoms are explicitly added, and the resulting molecular structures undergo a brief geometry optimization step using the Universal Force Field (UFF).

System construction (OpenMM). We then construct corresponding molecular systems based on these optimized geometries using `OpenMM`. The molecular interactions are parameterized via template generators provided by the `openmmforcefields` package. Unless stated otherwise, we employ the **OpenFF Sage 2.1.0** force field via the `SMIRNOFFTemplateGenerator`, utilizing the parameter file `openff-2.1.0.offxml`. Additionally, for comparison and ablation purposes, we optionally adopt the **GAFF-2.11** force field. Nonbonded interactions default to the `NoCutoff` method, which is suitable for small isolated molecular systems, though the Particle-Mesh Ewald (`PME`) method is also supported and can be used if required. Constraints are disabled, and the removal of center-of-mass (COM) motion is explicitly enabled at the system-building stage.

Unit conventions. Throughout the experiments, all public API interfaces strictly follow standard OpenMM unit conventions. Specifically, atomic positions $q \in \mathbb{R}^{N\times 3}$ are expressed in nanometers (nm); atomic momenta $p \in \mathbb{R}^{N\times 3}$ are given in units of dalton·nm/ps; energies, including kinetic (K), potential (U), and total Hamiltonian (H), are consistently reported in kilojoules per mole (kJ/mol). Atomic velocities are similarly represented in nm/ps.

Hamiltonian evaluation. We evaluate the total Hamiltonian energy as $H(q,p) = K(p) + U(q)$, utilizing a lightweight OpenMM `Context`. The potential energy $U(q)$ is directly obtained from the OpenMM context via the `getState()` function. Similarly, the kinetic energy $K(p)$ is computed from the context using `getKineticEnergy()` after atomic velocities are set according to their corresponding momenta and masses ($v = p/m$). Furthermore, each force term (e.g., bond, angle, torsion, nonbonded) in the system is assigned a unique force-group identifier, enabling individual extraction and analysis of potential energy contributions per term.

Impact Statement

This paper presents work whose goal is to advance the field of machine learning. There are many potential societal consequences of our work, none of which we feel must be specifically highlighted here.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The main claims are stated in the abstract and introduction and are supported by the theoretical and experimental results in the paper.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [N/A]

   Justification: The scope and practical constraints of the proposed approach are discussed in the manuscript and appendix.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best

judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: The assumptions and proofs for the theoretical results are provided in the paper and supplementary appendix.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: The paper describes the model, datasets, training, and evaluation settings in sufficient detail to support reproducibility.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [N/A]

   Justification: We provide detailed experimental and implementation descriptions in the submission, and we plan to release the code upon acceptance.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: The manuscript and appendix describe the experimental setting, evaluation protocol, and implementation details used in the study.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: The paper reports point estimates of trajectory-level metrics and runtime under a fixed evaluation protocol, but it does not report error bars, confidence intervals, or statistical significance tests for the main experimental claims.

   Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: The paper provides the relevant implementation and runtime information needed to understand the computational setup of the experiments.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We have reviewed the NeurIPS Code of Ethics and believe that the research conforms to it.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [N/A]

Justification: This work is a foundational methodological study on scientific simulation, and we do not identify a direct societal impact requiring a dedicated broader-impact discussion at submission time.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

Answer: [N/A]

Justification: This work does not involve releasing data or models with a high risk of misuse.

Guidelines:

- The answer [N/A] means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: Existing assets used in this work are credited in the manuscript, and their usage follows the corresponding terms and conventions.

Guidelines:

- The answer [N/A] means that the paper does not use existing assets.

- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: This submission does not release a new public asset alongside the anonymized paper submission.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: This work does not involve crowdsourcing or human-subject research.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

Justification: This work does not involve human-subject research and therefore IRB approval is not applicable.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: LLMs are not an important, original, or non-standard component of the core methodology in this work.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.