Bergische Universität Wuppertal

Fakultät für Mathematik und Naturwissenschaften

Institute of Mathematical Modelling, Analysis and Computational Mathematics (IMACM)

Preprint BUW-IMACM 26/05

Luca Di Persio, Matthias Ehrhardt, Youness Outaleb and Sofia Rizzotto

Port-Hamiltonian Neural Networks: From Theory to Simulation of Interconnected Stochastic Systems

(This is an updated/corrected version of Preprint 25/15)

April 15, 2026

http://www.imacm.uni-wuppertal.de

Port-Hamiltonian Neural Networks: From Theory to Simulation of Interconnected Stochastic Systems

Luca Di Persio[1], Matthias Ehrhardt[2], Youness Outaleb[3*], Sofia Rizzotto[1]

[1*]Department of Computer Science – College of Mathematics, University of Verona, Strada le Grazie 15, Verona, 37134, Italy.
[2]Chair of Applied and Computational Mathematics, University of Wuppertal, Gaußstrasse 20, Wuppertal, 42119, Germany.
[3]Doctoral School in Mathematics, University of Trento, Via Sommarive 14, Povo, 38123, Italy.

*Corresponding author(s). E-mail(s): youness.outaleb@unitn.it;
Contributing authors: luca.dipersio@univr.it;
ehrhardt@uni-wuppertal.de; sofia.rizzotto@studenti.univr.it;

Abstract

This work explores combining port-Hamiltonian structures with stochastic modeling and neural network learning. The revised framework retains Dirac structures in their standard fiberwise Courant form and introduces stochasticity via Stratonovich stochastic differential equations that evolve on that geometry. This separation of structure and dynamics results in a pathwise Stratonovich energy balance. After conversion to Itô form, the corrected generator identity makes the diffusion contribution explicit. Based on this, we formulate passivity in expectation through generator dissipativity rather than through deterministic pathwise inequalities. We also demonstrate how power-preserving interconnection preserves this notion under suitable diffusion assumptions. These results motivate the development of structure-aware stochastic port-Hamiltonian neural networks, parameterized to enforce skew-symmetry, positive semidefiniteness of the dissipation, and energy-based regularization. Numerical experiments on damped mass-spring, Duffing, and robotic systems demonstrate accurate long-term performance, and the stochastic theory provides energetic constraints to regularize the learned models.

Keywords: Stochastic port-Hamiltonian systems, port-Hamiltonian neural networks, passivity in expectation, interconnection, structure-preserving learning.

MSC Classification: 37N40 , 37B52 , 60G10 , 60H10 , 93C55

1 Introduction

This work examines the use of stochastic port-Hamiltonian systems in structure-preserving learning. Port-Hamiltonian models use Hamiltonians, skew-symmetric interconnection operators, and power ports to encode interconnection, storage, dissipation, and control. In the deterministic setting, their geometric backbone is provided by Dirac structures.

However, when stochastic forcing is introduced, it becomes necessary to separate the geometric structure from the stochastic calculus. In the revised framework, the Dirac structure remains fiberwise and time-independent; randomness enters only through Stratonovich stochastic differential equations that evolve on top of the structure. This approach avoids history-dependent definitions of orthogonality, makes the Itô correction in the energy balance explicit, and establishes a meaningful notion of passivity in expectation.

The same decomposition is useful for learning. We parameterize Hamiltonians, interconnection matrices, dissipation operators, port maps, and diffusion fields in a way that maintains the learned model's compatibility with the port-Hamiltonian form. The latter allows us to combine standard data-fitting objectives with losses derived from corrected Itô energy drift and generator-based passivity.

The paper is organized as follows: Section 2 reviews deterministic and discrete port-Hamiltonian preliminaries. Section 3 formulates the stochastic port-Hamiltonian dynamics and derives the Stratonovich and Itô energy balances. Section 4 develops passivity in expectation and its supermartingale characterization. Section 5 discusses interconnection, a representative stochastic agent model, and the learning implications of the theory. Section 6 presents the numerical experiments. Section 7 concludes.

2 Preliminaries on Port-Hamiltonian and Stochastic Models

2.1 Basic Concepts

Cordoni, Di Persio, and Muradore [3] describe an *input-state-output* (I-S-O) deterministic PHS using a geometric, coordinate-free formulation in terms of Poisson brackets:

$$\begin{cases} \dot{x} &= X_H(x) + \sum_{i=1}^{m} u_i X_{H_{g_i}}(x), \\ y_i &= \{H, H_{g_i}\}. \end{cases} \tag{1}$$

This is called an *(explicit) input-state-output port-Hamiltonian system* (PHS) on a Poisson manifold $(X, \{\cdot,\cdot\})$ with a Hamiltonian function $H \in C^\infty(\mathcal{X})$, $x \in \mathbb{R}^n$, the i-th input $u_i \in U$, the i-th output $y_i \in U^*$, and the Hamiltonian vector field $X_{H_{g_i}}$ associated with the Hamiltonian H_{g_i}. In local coordinates, the previous system (1)

reads

$$\begin{cases} \dot{x} & = J(x)\partial_x H(x) + \sum_{i=1}^{m} u_i g_i(x), \\ y_i & = g_i^\top(x)\partial_x H. \end{cases} \tag{2}$$

Moreover, given $X_H^L(\cdot) := [\cdot, H]_L$, we can define the *(explicit) input-state-output port-Hamiltonian system with dissipation* as follows:

$$\begin{cases} \dot{x} & = X_H^L(x) + \sum_{i=1}^{m} u_i H_{g_i}(x), \\ y_i & = [H, H_{g_i}], \end{cases} \tag{3}$$

and in local coordinates, this system (3) reads

$$\begin{cases} \dot{x} & = \big(J(x) - R(x)\big)\partial_x H(x) + \sum_{i=1}^{m} u_i g_i(x), \\ y_i & = g_i^\top(x)\partial_x H(x), \end{cases} \tag{4}$$

where $R(x) := (g^R(x))^\top \tilde{R}(x) g^R(x)$.

Consider a physical system consisting of energy-storing elements, energy-dissipating elements, and power ports. These elements are connected by power-preserving links that can only transfer energy and cannot generate it. Such a system can be described by extending the port-Hamiltonian system framework to implicit systems, i.e., systems with algebraic constraints. Given a state space $\mathcal{X}$ (a smooth manifold whose elements represent the energy stored in the system), a vector space of flow variables $\mathcal{V}$ and its dual space of effort variables $\mathcal{V}^*$ (representing the power ports), a geometric Dirac structure, $\mathcal{D}$, and a Hamiltonian function, $\mathcal{H}$, representing the total energy of the system in a given state, we can define an *implicit port-Hamiltonian system* corresponding to $(\mathcal{X}, \mathcal{V}, \mathcal{D}, \mathcal{H})$ as

$$v = -\dot{x} \quad \text{and} \quad v^* = \frac{\partial \mathcal{H}}{\partial x}(x)\,, \tag{5}$$

implying the system is defined by

$$\Big(-\dot{x}, \frac{\partial H}{\partial x}(x), f, e\Big) \in \mathcal{D}(x). \tag{6}$$

The Dirac structure discussed here describes the internal interconnection behavior of a port-Hamiltonian system and provides a mathematical framework for understanding how its components interact. A key attribute of Dirac structures is that their power-conserving composition yields another valid Dirac structure. This property is crucial because it implies that any interconnection of port-Hamiltonian systems that preserves power will also result in a valid system that maintains the underlying principle of energy conservation. Specifically, when combining these systems, the overall Dirac structure is constructed by integrating the individual structures, and the total

Hamiltonian of the interconnected system is expressed as the sum of the Hamiltonians of its components.

Furthermore, a fundamental characteristic of these systems is the concept of *passivity*, which states that the total energy supplied to the system must equal or exceed the energy released, assuming no losses due to friction, resistance, or other dissipative effects. This principle of passivity is essential for ensuring stability in various control applications because it reveals how a system behaves in response to energy inputs and outputs. Passivity naturally arises from the underlying Dirac structure and is closely tied to the energy-dissipation relationship inherent to port-Hamiltonian systems. This relationship guarantees that energy is neither created nor destroyed within the system; it is only shifted between different forms and components. Ultimately, this relationship enables robust, stable control strategies.

In summary, understanding the dynamics and stability of PHSs require grasping the interplay between the Dirac structure, power conservation, and passivity. This interplay allows us to design and analyze complex, interconnected systems for various engineering and physical applications.

2.2 Discrete Systems

Following the algebraic viewpoint used in structure-preserving discretizations [9, 12], we formulate the discrete theory in finite-dimensional terms. Let V_h be a finite-dimensional real vector space of discrete flows and let V_h^* denote its dual space of discrete efforts. The discrete analog of the Courant pairing is the symmetric bilinear form

$$\big\langle\!\big\langle v \oplus \alpha,\, w \oplus \beta \big\rangle\!\big\rangle := \langle \alpha,\, w\rangle + \langle \beta,\, v\rangle. \tag{7}$$

A discrete Dirac structure is a linear subspace $D_h \subset V_h \oplus V_h^*$ that is maximally isotropic with respect to (7).

If two subsystems with flow spaces V_A and V_B are interconnected through a power-preserving constraint on an interconnection space V_I, then the external relation is obtained by composing the corresponding maximally isotropic subspaces, so that we obtain the finite-dimensional analog of composing Dirac structures and preserving maximal isotropy.

Definition 1 (Discrete port-Hamiltonian system) Let $H_h\colon V_h \to \mathbb{R}$ be a discrete Hamiltonian, $J_h\colon V_h^* \to V_h$ skew-symmetric, $R_h\colon V_h^* \to V_h$ symmetric positive semidefinite, and $G_h\colon \mathbb{R}^p \to V_h$ a port map. A discrete port-Hamiltonian system is defined by

$$\dot{x}_h = \big[J_h - R_h\big]\nabla H_h(x_h) + G_h u,\; y_h = G_h^* \nabla H_h(x_h). \tag{8}$$

Its energy identity reads

$$\frac{\mathrm{d}}{\mathrm{d}t} H_h(x_h) = -\langle \nabla H_h(x_h),\, R_h \nabla H_h(x_h)\rangle + \langle y_h,\, u\rangle. \tag{9}$$

This formulation is sufficient for the later numerical discussion and avoids introducing undefined spaces of discrete vector fields and discrete one-forms.

3 Stochastic port-Hamiltonian systems

We separate structure from stochasticity. The geometric structure is a local fiberwise constraint on $TM \oplus T^*M$. Stochasticity enters only through stochastic differential equations evolving on top of that structure.

Dirac structures are defined fiberwise and are independent of a driving signal. Stochastic calculus becomes relevant only after the dynamics are chosen. We use Stratonovich calculus as the geometric representation because it satisfies the classical chain rule, and we use Itô calculus for expectation-based balance laws and generator estimates.

3.1 Dirac structures and port interconnection

Let M be a smooth manifold and set $\mathbb{T}M := TM \oplus T^*M$. For each $x \in M$ and $v, w \in T_xM$, $\alpha, \beta \in T_x^*M$, define the symmetric Courant pairing

$$\big\langle\!\big\langle v \oplus \alpha,\, w \oplus \beta \big\rangle\!\big\rangle_x := \langle \alpha,\, w \rangle + \langle \beta,\, v \rangle. \tag{10}$$

Definition 2 (Dirac structure) A Dirac structure on M is a smooth vector subbundle $D \subset \mathbb{T}M$ such that each fiber D_x is maximally isotropic with respect to (10).

For ports, we consider flow-effort pairs $f, e \in \mathbb{R}^p$ equipped with the symmetric pairing

$$\big\langle\!\big\langle f \oplus e,\, f' \oplus e' \big\rangle\!\big\rangle := e^\top f' + (e')^\top f. \tag{11}$$

A power-preserving interconnection constraint is a maximally isotropic subspace of $\mathbb{R}^p \oplus \mathbb{R}^p$ with respect to (11).

In local coordinates with $M = \mathbb{R}^n$, a skew-symmetric interconnection matrix $J(x) = -J(x)^\top$ and a port map $G(x)$ define the port relation

$$\dot{x} = J(x)e_x + G(x)f, \quad e = -G(x)^\top e_x, \tag{12}$$

whose graph is maximally isotropic with respect to the direct-sum pairing induced by (10) and (11). In the remainder of the paper, we work in these local coordinates.

3.2 Stochastic port-Hamiltonian dynamics

Let W be an m-dimensional Brownian motion on a filtered probability space satisfying the usual conditions. Let $H\colon \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable, $J\colon \mathbb{R}^n \to \mathbb{R}^{n\times n}$ skew-symmetric, $R\colon \mathbb{R}^n \to \mathbb{R}^{n\times n}$ symmetric positive semidefinite, and $G\colon \mathbb{R}^n \to \mathbb{R}^{n\times p}$ a port map. Let $u_t \in \mathbb{R}^p$ be an adapted input and let $\sigma_k\colon \mathbb{R}^n \to \mathbb{R}^n$ denote diffusion vector fields.

The stochastic port-Hamiltonian dynamics is written in Stratonovich form as

$$\delta X_t = \Big[\big(J(X_t) - R(X_t)\big)\nabla H(X_t) + G(X_t)u_t\Big]\delta t + \sum_{k=1}^{m} \sigma_k(X_t)\,\delta W_t^k, \tag{13}$$

with output

$$y_t := G(X_t)^\top \nabla H(X_t). \tag{14}$$

Proposition 1 (Stratonovich energy balance) *Let X solve (13). Then*

$$\delta H(X_t) = -\nabla H(X_t)^\top R(X_t) \nabla H(X_t)\, \delta t + y_t^\top u_t\, \delta t + \sum_{k=1}^{m} z_k(X_t)\, \delta W_t^k, \tag{15}$$

where $z_k(x) := \nabla H(x)^\top \sigma_k(x)$.

Proof The Stratonovich chain rule gives $\delta H(X_t) = \nabla H(X_t)^\top \delta X_t$. Substituting (13) yields three contributions. The interconnection term vanishes because $a^\top J(x) a = 0$ for all $a \in \mathbb{R}^n$, the dissipation term contributes $-\nabla H^\top R \nabla H$, and the port term equals $y_t^\top u_t$, and we obtain (15). □

The equivalent Itô state equation is

$$\mathrm{d}X_t = b(X_t, u_t)\, \mathrm{d}t + \sum_{k=1}^{m} \sigma_k(X_t)\, \mathrm{d}W_t^k, \tag{16}$$

with drift

$$b(x,u) = \big(J(x) - R(x)\big)\nabla H(x) + G(x)u + \frac{1}{2}\sum_{k=1}^{m} D\sigma_k(x)\, \sigma_k(x), \tag{17}$$

where $D\sigma_k(x)$ denotes the Jacobian of σ_k.

Proposition 2 (Itô energy balance and generator) *Along solutions of (16), the Hamiltonian satisfies*

$$\mathrm{d}H(X_t) = \mathcal{L}^{u_t} H(X_t)\, \mathrm{d}t + \sum_{k=1}^{m} z_k(X_t)\, \mathrm{d}W_t^k, \tag{18}$$

where

$$\mathcal{L}^u H(x) := \nabla H(x)^\top b(x,u) + \frac{1}{2}\sum_{k=1}^{m} \sigma_k(x)^\top \nabla^2 H(x)\, \sigma_k(x). \tag{19}$$

Equivalently,

$$\mathcal{L}^u H(x) = -\nabla H(x)^\top R(x) \nabla H(x) + y(x)^\top u + \frac{1}{2}\sum_{k=1}^{m} Dz_k(x)\, \sigma_k(x), \tag{20}$$

with $y(x) := G(x)^\top \nabla H(x)$.

Proof Apply Itô's formula to $H(X_t)$ along (16). Using the explicit drift (17) and the skew-symmetry of J yields (19). Since $z_k(x) = \nabla H(x)^\top \sigma_k(x)$, the correction terms can be regrouped as

$$\nabla H(x)^\top D\sigma_k(x)\, \sigma_k(x) + \sigma_k(x)^\top \nabla^2 H(x)\, \sigma_k(x) = Dz_k(x)\, \sigma_k(x),$$

which gives (20). □

3.3 Generator-based dissipativity

The Itô correction in (20) makes the diffusion contribution explicit and is the correct starting point for expectation-based passivity.

Theorem 3 (Generator-based dissipativity) *Assume that for all* $x \in \mathbb{R}^n$

$$\frac{1}{2}\sum_{k=1}^{m} Dz_k(x)\,\sigma_k(x) \le \nabla H(x)^\top R(x)\nabla H(x),\ z_k(x) := \nabla H(x)^\top \sigma_k(x). \tag{21}$$

Then, for every admissible input u,

$$\mathbb{E}\big[H(X_t)\big] - H(X_0) \le \mathbb{E}\int_0^t y_s^\top u_s\,\mathrm{d}s. \tag{22}$$

Proof Taking expectations in (18) eliminates the martingale term. Combining (20) with the pointwise condition (21) yields $\mathcal{L}^u H(x) \le y(x)^\top u$. Integrating in time gives (22). □

4 Passivity in Stochastic Systems

For Brownian-driven systems, deterministic almost-sure energy bounds are generally destroyed by martingale fluctuations. The relevant notion is therefore passivity in expectation, equivalently, a generator inequality.

Definition 3 (Weak passivity and limited-scope strong passivity) Let X solve (16) and let H be a nonnegative storage function. The system is weakly passive with supply rate $y^\top u$ if for every $t \ge 0$ and every admissible input u one has

$$\mathbb{E}\big[H(X_t)\big] - H(X_0) \le \mathbb{E}\int_0^t y_s^\top u_s\,\mathrm{d}s. \tag{23}$$

Equivalently, weak passivity holds if $\mathcal{L}^u H(x) \le y(x)^\top u$ pointwise for all $x \in \mathbb{R}^n$ and all $u \in \mathbb{R}^p$. Strong passivity is considered only when the perturbation has finite variation, so that the martingale term in (18) is absent. In that case, the pathwise inequality follows by integrating (15).

Proposition 4 (Supermartingale characterization) *If* $\mathcal{L}^u H(x) \le y(x)^\top u$ *for all* x *and* u, *then*

$$H(X_t) - \int_0^t y_s^\top u_s\,\mathrm{d}s \tag{24}$$

is a supermartingale. In particular, (23) *holds.*

Proof Using (18),

$$H(X_t) - \int_0^t y_s^\top u_s\,\mathrm{d}s = H(X_0) + \int_0^t \big[\mathcal{L}^{u_s} H(X_s) - y_s^\top u_s\big]\,\mathrm{d}s + \sum_{k=1}^{m}\int_0^t z_k(X_s)\,\mathrm{d}W_s^k.$$

The drift is nonpositive by assumption, and the stochastic integral is a martingale under the standard integrability conditions. Hence, the process in (24) is a supermartingale. □

Lemma 1 (Strong passivity typically fails under Brownian forcing) Assume that the energy balance contains a martingale term of the form $M_t = \int_0^t \gamma_s \, \mathrm{d}W_s$ with predictable γ such that $\int_0^T \gamma_s^2 \, \mathrm{d}s > 0$ with positive probability for some $T > 0$. Then for every constant $C \in \mathbb{R}$ one has

$$\mathbb{P}\Big[\sup_{t\in[0,T]} M_t > C\Big] > 0. \tag{25}$$

Consequently, unless the diffusion is degenerate in the sense that $\gamma \equiv 0$, one cannot expect deterministic pathwise dissipation inequalities to hold on all sample paths.

Proof By the Dambis-Dubins-Schwarz theorem, M is a time-changed Brownian motion on the event where its quadratic variation is positive. The supremum of Brownian motion exceeds any prescribed constant with positive probability, which yields the claim. □

Remark 1 Lemma 1 shows that deterministic pathwise dissipation inequalities are unavailable for generic Brownian-driven systems unless the diffusion contribution to the energy vanishes, which is the reason why the manuscript formulates passivity through (23) and the generator inequality rather than through almost-sure Stratonovich bounds.

5 Application and Generalization

5.1 Interconnection of Multiple SPHS

One of the key advantages of the port-Hamiltonian viewpoint is that interconnected systems inherit an energy-based structure. In the stochastic setting, the relevant stability notion is weak passivity in expectation, so the interconnection result must track both supplied power and diffusion. For notational simplicity, we state the result for two subsystems. The extension to finitely many modules is immediate. Figure 1 illustrates the corresponding multi-module picture.

Theorem 5 (Interconnection preserves weak passivity) *Consider two Itô stochastic port-Hamiltonian systems on $\mathbb{R}^{n_1}$ and $\mathbb{R}^{n_2}$ of the form (16) with Hamiltonians H_1 and H_2, outputs y_1 and y_2, inputs u_1 and u_2, and generators $\mathcal{L}_1$ and $\mathcal{L}_2$. Assume each subsystem is weakly passive, meaning*

$$\mathcal{L}_1^{u_1} H_1(x_1) \le y_1^\top u_1, \;\; \mathcal{L}_2^{u_2} H_2(x_2) \le y_2^\top u_2.$$

Assume the interconnection is power preserving in the sense that there exist external port variables u and y such that

$$y_1^\top u_1 + y_2^\top u_2 = y^\top u. \tag{26}$$

Assume moreover that the diffusion matrix of the interconnected system is block diagonal, or more generally that the cross-quadratic variations between the two state components vanish. Then the interconnected system with state $X = (X_1, X_2)$ and storage $H := H_1 + H_2$ is weakly passive with respect to the external supply $y^\top u$.

Proof Under the vanishing cross-variation assumption, the generator of the interconnected system acts additively on separable observables. Hence

$$\mathcal{L}^u H(X) = \mathcal{L}_1^{u_1} H_1(X_1) + \mathcal{L}_2^{u_2} H_2(X_2) \leq y_1^\top u_1 + y_2^\top u_2.$$

Using the power-preserving relation (26) gives $\mathcal{L}^u H(X) \leq y^\top u$, which is precisely weak passivity for the interconnected system. □

Remark 2 If correlated noise introduces mixed quadratic variation terms, the generator acquires additional mixed Hessian contributions. In that case, the same argument applies once those mixed terms are controlled by a suitable nonpositive bound.

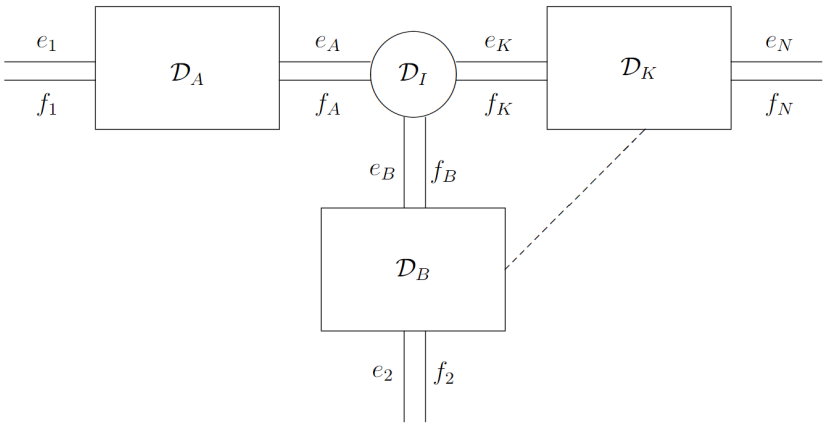


Fig. 1: Interconnection of N implicit port-Hamiltonian systems.

5.2 Discrete Stochastic PHS

Consider the continuous stochastic port-Hamiltonian system

$$\begin{cases} dX_t &= \big((J-R)\partial_x H(X_t) + g(X_t)u_t\big)\,dt + \xi(X_t)\,\delta W_t, \\ y_t &= g^\top(X_t)\partial_x H(X_t), \\ z_t &= \xi^\top(X_t)\partial_x H(X_t), \end{cases} \tag{27}$$

where $J = -J^\top$, R positive semidefinite, g represents control port, H is the Hamiltonian, $u \in U$ is the control input, $y \in Y$ is the output of the system, ξ is a matrix, z is the associated effort to δW_t and W is a standard Brownian motion adapted to the reference filtration $(\mathcal{F}_t)_{t\geq 0}$. Then we can introduce the discretization

$$\dot{X}(t_0^k + \tau h) = -f(t_0^k + \tau h) = -\sum_{j=1}^{s} f_j^k\, l_j(\tau), \tag{28}$$

with

$$\dot{X}(t_i^k) := -f_i^k, \quad l_i(\tau) = \prod_{j=1}^{s} \frac{\tau - c_j}{c_i - c_j}, \quad \tau \in [0,1],$$

where l_i is the i^{th} Lagrange interpolation polynomial of order s and τ is the normalized time parameterizing the sampling intervals. Thus, we can generalize the continuous SPHS to a discrete form, preserving the structure of the Hamiltonian and the control inputs, as follows:

Definition 4 (Discrete stochastic port-Hamiltonian system, [9]) A *discrete stochastic port-Hamiltonian system* can be written as

$$\begin{cases} X(t_0^k + c_i h) & = x_0^k - h \sum_{i=1}^{s} a_{ij} f_j^k, \\ X(t_0^k + h) & = x_0^k - h \sum_{i=1}^{s} a_j f_i^k, \\ -a_{ij} f^k & = (J_j^k - R_j^k)\, a_{ij} e_j^k + a_{ij} g_j^k u_j^k + b_{ij} \xi_j^k \, \Delta W, \end{cases} \tag{29}$$

where ΔW is a truncated centered Gaussian random variable with variance h, $a_{ij} = \int_0^{c_i} l_j(\sigma)\, d\sigma$, $a_j = \int_0^1 l_j(\sigma)\, d\sigma$ and $M = M^\top$, cf. [9].

Note that in the discrete case, the system is *passive* if it holds

$$\mathbb{E}\big[\Delta H^k\big] \leq h\, \mathbb{E}\big[(y^k)^\top u^k\big]. \tag{30}$$

5.3 Stochastic Motion Model of Agents

Ehrhardt, Kruse, and Tordeux present an application of a stochastic motion model of agents [6], analyzing the case in which agents' positions and velocities are modeled as a ring structure. The initial positions and velocities are set, and the system is governed by differential equations involving the velocities and positions of neighboring agents [6]. The dynamic equations of the agents are given by:

$$\begin{cases} dQ_n(t) & = \big(p_{n+1}(t) - p_n(t)\big)\, dt, \\ dp_n(t) & = \big(U'(Q_n(t)) - U'(Q_{n-q}(t))\big)\, dt \\ & \quad + \beta\big(p_{n+1}(t) - 2p_n(t) + p_{n-1}(t)\big)\, dt + \sigma\, dW_n(t), \end{cases} \tag{31}$$

with $Q(0) = Q_0 \in [0, +\infty)^N$ the initial distance, $p(0) = p_0$ the initial velocity, $\beta \in (0, +\infty)$ a dissipation rate, $\sigma \in \mathbb{R}$ the noise volatility, U' the derivative of a convex potential $U \in C^1(\mathbb{R}, [0, +\infty))$ and $W = (W_n)_{n=1}^N \colon [0, +\infty) \times \Omega \to \mathbb{R}^N$ an N-dimensional standard Brownian motion defined on $(\Omega, \mathcal{F}, \mathbb{P})$. These equations represent the acceleration of the n^{th} agent depending on its neighbors' velocities and the Brownian motion's stochastic perturbations [7].

The motion of the agents is further formulated using a stochastic port-Hamiltonian framework

$$dZ(t) = (J - R)\nabla H(Z(t))\, dt + G\, dW(t),$$

where $Z(t) = \big(Q(t), p(t)\big)^\top \in \mathbb{R}^{2N}$, $t \in [0, +\infty)$, J and R are defined as skew-symmetric and symmetric positive semidefinite matrices, respectively. This formulation allows the application of Hamiltonian dynamics to model agents' behavior under stochastic influences. In particular, the Hamiltonian is independent of Q and U, and its expectation could increase over time. Moreover, describing the limiting behavior of these stochastic systems is challenging, so the authors [6] focus on the specific scenario where the quadratic function characterizes the potential

$$U(x) = \frac{(\alpha x)^2}{2} \quad x \in \mathbb{R},\ \alpha \in (0, \infty), \tag{32}$$

in which the process reads

$$dZ(t) = BZ(t)\,dt + G\,dW(t), \quad Z(0) = (Q_0, p_0)^\top, \tag{33}$$

where B is defined such that $BZ(t) = (J - R)\nabla H(Z(t))$. In this case, the resulting process converges in distribution to a normal distribution with known expectation and covariance matrix as $t \to \infty$.

5.4 Structure-preserving learning for stochastic port-Hamiltonian models

The corrected theory in Sections 3–4 can be made operational in learning by constraining the model class itself. Assume data consist of state samples X, control inputs u, and, when available, stochastic increments. We learn an Itô model compatible with (16) rather than an unconstrained vector field.

Introduce neural parameterizations H_θ, J_θ, R_θ, G_θ, and $\sigma_{k,\theta}$ such that

$$\begin{aligned} b_\theta(x,u) &= \big(J_\theta(x) - R_\theta(x)\big)\nabla H_\theta(x) + G_\theta(x)u + \frac{1}{2}\sum_{k=1}^{m} D\sigma_{k,\theta}(x)\,\sigma_{k,\theta}(x), \\ \sigma_\theta(x) &:= \big[\sigma_{1,\theta}(x)\ \cdots\ \sigma_{m,\theta}(x)\big]. \end{aligned} \tag{34}$$

The constraints $J_\theta(x) = -J_\theta(x)^\top$ and $R_\theta(x) \succeq 0$ are enforced by construction via

$$J_\theta(x) := A_\theta(x) - A_\theta(x)^\top,\ R_\theta(x) := B_\theta(x)B_\theta(x)^\top,$$

where A_θ and B_θ are unrestricted neural outputs. The corresponding predicted energy drift is

$$\mu_\theta(x,u) := \mathcal{L}^u_\theta H_\theta(x) = \nabla H_\theta(x)^\top b_\theta(x,u) + \frac{1}{2}\sum_{k=1}^{m} \sigma_{k,\theta}(x)^\top \nabla^2 H_\theta(x)\,\sigma_{k,\theta}(x). \tag{35}$$

A generic structure-aware training objective may then combine one-step consistency with an Itô energy residual:

$$\begin{aligned}\mathcal{L}_{\text{bridge}}(\theta) = \frac{1}{N}\sum_{j=1}^{N}\left\|X_{j+1} - X_j - b_\theta(X_j,u_j)\Delta t - \sigma_\theta(X_j)\Delta W_j\right\|^2 \\ + \lambda\,\frac{1}{N}\sum_{j=1}^{N}\left|H_\theta(X_{j+1}) - H_\theta(X_j) - \mu_\theta(X_j,u_j)\Delta t\right|^2\end{aligned} \tag{36}$$

The second term measures compliance with the corrected stochastic energy balance, while additional penalties can enforce generator passivity.

6 Applications of Port-Hamiltonian Neural Networks

In this section, we evaluate learning architectures that implement the stochastic port-Hamiltonian model introduced in Sections 3–5. The learned drift is constrained to the form (34). When the data are deterministic, the diffusion terms are set to zero, and when stochastic forcing is present, the diffusion is learned from the data. The training objectives combine data-fidelity terms with structure terms derived from the corrected Itô energy balance and generator-based passivity.

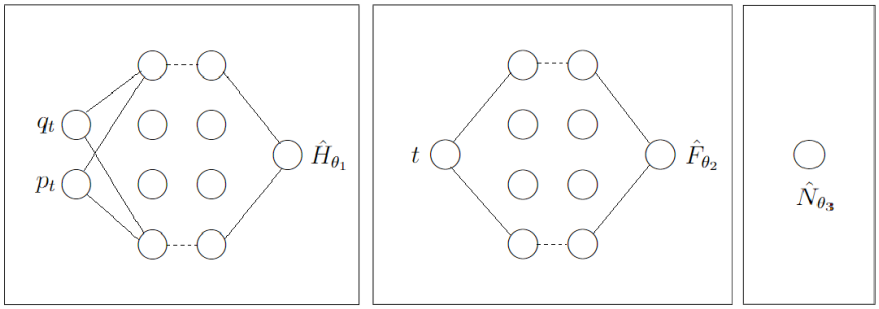


Fig. 2: Schematic representation of a port-Hamiltonian neural network (pHNN).

Figure 2 illustrates the architectural philosophy behind the pHNN models used below: the network learns Hamiltonian, interconnection, dissipation, port, and diffusion terms under algebraic constraints rather than learning an unconstrained vector field. The following subsections specify the parameterization and the losses used in the experiments.

6.1 Structure-constrained stochastic port-Hamiltonian model

We use Euclidean special cases of (34). A neural network defines a Hamiltonian $H_\theta\colon \mathbb{R}^n \to \mathbb{R}$. Another network defines $A_\theta(x) \in \mathbb{R}^{n\times n}$ and we set

$$J_\theta(x) := A_\theta(x) - A_\theta(x)^\top. \tag{37}$$

A network defines $B_\theta(x)$ and we set

$$R_\theta(x) := B_\theta(x)B_\theta(x)^\top. \tag{38}$$

A port map $G_\theta(x) \in \mathbb{R}^{n\times p}$ gives the output

$$y_\theta(x) := G_\theta(x)^\top \nabla H_\theta(x). \tag{39}$$

Diffusion is modeled through vector fields $\sigma_{k,\theta}(x)$, collected in

$$\sigma_\theta(x) := \big[\sigma_{1,\theta}(x) \ \cdots \ \sigma_{m,\theta}(x)\big].$$

The corresponding Itô drift is

$$b_\theta(x,u) = \big(J_\theta(x) - R_\theta(x)\big)\nabla H_\theta(x) + G_\theta(x)u + \frac{1}{2}\sum_{k=1}^{m} D\sigma_{k,\theta}(x)\,\sigma_{k,\theta}(x). \tag{40}$$

By construction, J_θ is skew-symmetric and R_θ is positive semidefinite, so the learned model retains the algebraic port-Hamiltonian constraints.

6.2 Losses derived from the stochastic port-Hamiltonian model

Assume a dataset of samples $\{(X_j, u_j)\}_{j=0}^N$ on a uniform grid with step size Δt and, when stochastic forcing is present, corresponding Brownian increments ΔW_j. The learned model is the Itô SDE (16) with coefficients $(b_\theta, \sigma_\theta)$.

A first loss enforces one-step consistency of the state increments:

$$\mathcal{L}_{\mathrm{dyn}}(\theta) := \frac{1}{N}\sum_{j=0}^{N-1} \big\| X_{j+1} - X_j - b_\theta(X_j,u_j)\Delta t - \sigma_\theta(X_j)\Delta W_j \big\|_2^2 \tag{41}$$

Define the predicted energy drift by

$$\begin{aligned}\mu_\theta(X_j,u_j) &:= \nabla H_\theta(X_j)^\top b_\theta(X_j,u_j) + \frac{1}{2}\sum_{k=1}^{m} \sigma_{k,\theta}(X_j)^\top \nabla^2 H_\theta(X_j)\,\sigma_{k,\theta}(X_j) \\ &= \mathcal{L}_\theta^{u_j} H_\theta(X_j).\end{aligned} \tag{42}$$

A second loss enforces the corrected one-step Itô energy balance. Define

$$z_{k,\theta}(x) := \nabla H_\theta(x)^\top \sigma_{k,\theta}(x). \tag{43}$$

When the Brownian increments ΔW_j are available, we set

$$\mathcal{L}_{\mathrm{EB}}(\theta) := \frac{1}{N}\sum_{j=0}^{N-1}\Big|H_\theta(X_{j+1}) - H_\theta(X_j) - \mu_\theta(X_j,u_j)\Delta t - \sum_{k=1}^{m} z_{k,\theta}(X_j)\Delta W_j^k\Big|^2. \quad (44)$$

When ΔW_j are not observed, (44) is replaced by a conditional-mean version based on local averaging or transition likelihood; in that case, the martingale increment is not penalized sample-wise.

A third loss penalizes violations of generator passivity:

$$\mathcal{L}_{\mathrm{pass}}(\theta) := \frac{1}{N}\sum_{j=0}^{N-1}\Big[\max\big\{0,\ \mu_\theta(X_j,u_j) - y_\theta(X_j)^\top u_j\big\}\Big]^2. \quad (45)$$

The combined structural objective is

$$\mathcal{L}_{\mathrm{struct}}(\theta) := \mathcal{L}_{\mathrm{dyn}}(\theta) + \lambda_{\mathrm{EB}}\mathcal{L}_{\mathrm{EB}}(\theta) + \lambda_{\mathrm{pass}}\mathcal{L}_{\mathrm{pass}}(\theta). \quad (46)$$

When only time derivatives $(\dot q_t, \dot p_t)$ are observed, we use the deterministic derivative-matching surrogate

$$\mathcal{L}_{qp} = ||\hat{\dot q}_t - \dot q_t||_2^2 + ||\hat{\dot p}_t - \dot p_t||_2^2. \quad (47)$$

The latter is the noise-free analog of (41) and remains useful for the baseline comparison below.

6.3 Diffusion calibration via moment matching

The structural losses above constrain energy drift and passivity, but they do not by themselves match the covariance of one-step stochastic increments. When stochastic increments are available, we therefore add the following diffusion-calibration term. Recall that the Itô SDE

$$\mathrm{d}X_t = f(X_t)\,\mathrm{d}t + \Sigma(X_t)\,\mathrm{d}W_t, \quad X_t = [q_t^\top, p_t^\top]^\top,$$

with Brownian motion W_t, satisfies over one sampling interval Δt the approximations $E[\Delta X_t] = f(X_t)\Delta t + O(\Delta t^2)$ and $\mathrm{cov}[\Delta X_t] = \Sigma\Sigma^\top(X_t)\Delta t + O(\Delta t^2)$. Substituting neural predictions $\hat f_\theta$ and $\hat\Sigma_\theta$ for the unknown drift and diffusion yields

$$\mathcal{L}_{\mathrm{DDpHNN}} = \frac{1}{\Delta t}\|\Delta X_t - \hat f_\theta \Delta t\|_2^2 + \frac{1}{\Delta t}\|\Delta X_t \Delta X_t^\top - \hat\Sigma_\theta\hat\Sigma_\theta^\top \Delta t\|_F^2, \quad (48)$$

with the decomposition

$$\mathcal{L}_{\mathrm{DDpHNN}} = \underbrace{\frac{1}{\Delta t}\|\Delta X_t - \hat f_\theta \Delta t\|_2^2}_{\mathcal{L}_{\mathrm{drift}}} + \underbrace{\frac{1}{\Delta t}\|\Delta X_t \Delta X_t^\top - \hat\Sigma_\theta\hat\Sigma_\theta^\top \Delta t\|_F^2}_{\mathcal{L}_{\mathrm{diff}}}.$$

We scale the diffusion term by λ to rebalance the training gradients and prevent the model from exaggerating the predicted noise magnitude, while leaving drift learning unaffected, hence providing

$$\mathcal{L}^{(\lambda)}_{\text{DDpHNN}} = \mathcal{L}_{\text{drift}} + \lambda\,\mathcal{L}_{\text{diff}}, \tag{49}$$

and to prevent unbounded growth of Σ_θ we add an ℓ_1 penalty ($\alpha \ll 1$):

$$\mathcal{L}^{(\lambda,\alpha)}_{\text{DDpHNN}} = \mathcal{L}_{\text{drift}} + \lambda\,\mathcal{L}_{\text{diff}} + \alpha\|\hat{\Sigma}_\theta\|_1. \tag{50}$$

This term complements, rather than replaces, the generator-based losses in Section 6.2: its role is to calibrate the stochastic covariance structure while $\mathcal{L}_{\text{EB}}$ and $\mathcal{L}_{\text{pass}}$ enforce the corrected energetic constraints.

6.4 Additional deterministic regularizers

For deterministic benchmarks such as the 3-DOF robotic arm and the Duffing oscillator, we also consider an auxiliary objective that combines local derivative fitting, centered energy regression, and a symplectic rollout penalty. et a minibatch $B = \{(x_i, f_i, h_i)\}_{i=1}^m$ be given, with states $x_i = (q_i, p_i)$, ground-truth vector field $f_i = (\dot{q}_i, \dot{p}_i)$, and ground-truth energy $h_i = H_{\text{true}}(x_i)$.

In our model, we define $H_\theta(x)$ and the Hamiltonian vector field

$$f_\theta(x) = J\nabla_x H_\theta(x) = \Big(\frac{\partial H_\theta}{\partial p},\ -\frac{\partial H_\theta}{\partial q}\Big), \quad J = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}.$$

We build the auxiliary objective from the following three terms:

Vector field MSE

$$\mathcal{L}_{\text{vf}} = \frac{1}{m}\sum_{i=1}^{m} \big\|f_\theta(x_i) - f_i\big\|_2^2.$$

This term predicts the Hamiltonian vector field $(\dot{q}, \dot{p}) = \big(\frac{\partial H}{\partial p}, -\frac{\partial H}{\partial q}\big)$ and measures its discrepancy from the observed derivatives.

Batch centered energy MSE

We regress the scalar energy $H(x)$, but we compare centered values $H - \bar{H}_{\text{batch}}$, so the loss is invariant to an additive constant. In Hamiltonian mechanics, the equations of motion depend on gradients of H, so adding a constant to H changes nothing physically. Centering, therefore, teaches the network the shape and scale of the energy landscape without being sensitive to an arbitrary offset.

$$\bar{H}_\theta = \frac{1}{m}\sum_{i=1}^{m} H_\theta(x_i), \qquad \bar{h} = \frac{1}{m}\sum_{i=1}^{m} h_i,$$

$$\mathcal{L}_{\mathrm{E}} = \frac{1}{m}\sum_{i=1}^{m}\Big(\big(H_\theta(x_i) - \bar{H}_\theta\big) - \big(h_i - \bar{h}\big)\Big)^2.$$

One period symplectic rollout MSE

For a set of K one-period rollouts $\{X_t^{(r)}\}_{t=0}^{L}$ with step Δt (so $T = L\Delta t$), generate a predicted path $\{\hat{X}_t^{(r)}\}$ with the implicit midpoint update

$$\hat{X}_{t+1}^{(r)} = \hat{X}_t^{(r)} + \Delta t\, J\nabla H_\theta\Big(\tfrac{\hat{X}_t^{(r)} + \hat{X}_{t+1}^{(r)}}{2}\Big), \quad \hat{X}_0^{(r)} = X_0^{(r)},$$

and define

$$\mathcal{L}_{\mathrm{roll}} = \frac{1}{K}\sum_{r=1}^{K}\frac{1}{L+1}\sum_{t=0}^{L}\big\|\hat{X}_t^{(r)} - X_t^{(r)}\big\|_2^2.$$

$\mathcal{L}_{\mathrm{roll}}$ enforces long-horizon trajectory fidelity. The resulting auxiliary deterministic objective is

$$\mathcal{L}_{VER} = \mathcal{L}_{\mathrm{vf}} + \lambda_{\mathrm{E}}\,\mathcal{L}_{\mathrm{E}} + \mathcal{L}_{\mathrm{roll}}, \tag{51}$$

with $\lambda_{\mathrm{E}} > 0$ a tuning weight. In practice, this combination is useful because it simultaneously constrains local derivatives, the learned energy landscape, and long-horizon rollout fidelity under a structure-preserving integrator. We now test the resulting objectives. The learning curves, phase portraits, energy traces, state trajectories, and error statistics are presented in the following figures and discussed below.

6.5 Tests and Results

6.5.1 Comparing $\mathcal{L}_{qp}$ and $\mathcal{L}_{DDpHNN}$

We test the first and second objectives on the canonical damped-mass-spring system.

Training loss

Figure 3 confirms stable convergence for both objectives. The loss $\mathcal{L}_{\mathrm{DDpHNN}}$ (orange) decays one order of magnitude below the deterministic loss $\mathcal{L}_{\mathrm{qp}}$ (blue) and plateaus without signs of overfitting, showing that the additional diffusion term is well behaved.

Phase portrait and energy evolution

As shown in Figure 4a, both models closely follow the true spiral trajectory for the entire time window. Figure 4b tracks the Hamiltonian $H(q, p)$. Both networks exhibit the dissipative trend.

Position and momentum

All three trajectories begin with the same damped oscillatory pattern and gradually diverge as the rollout progresses, demonstrating a consistent decrease in amplitude.

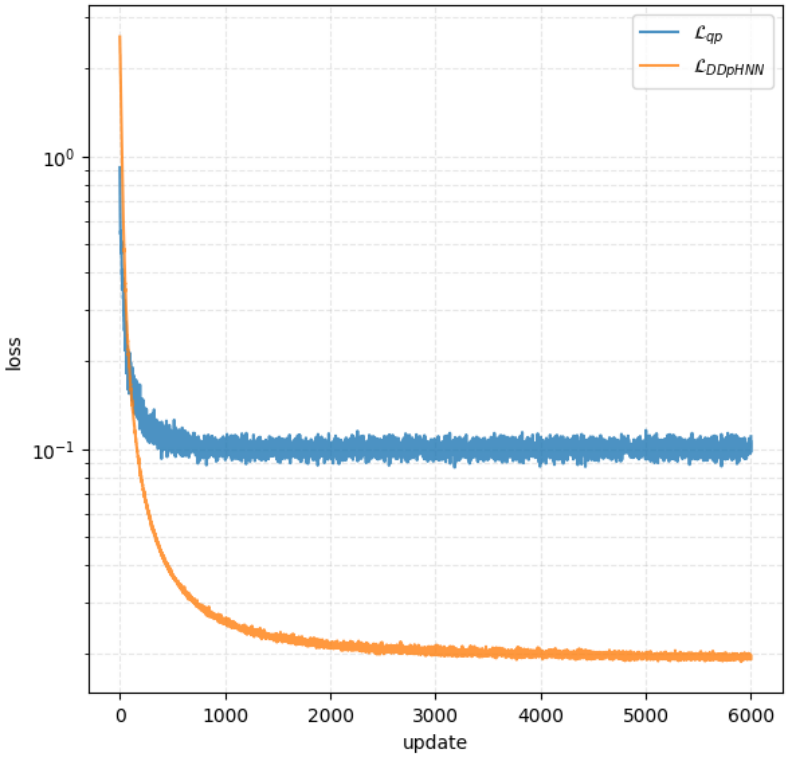
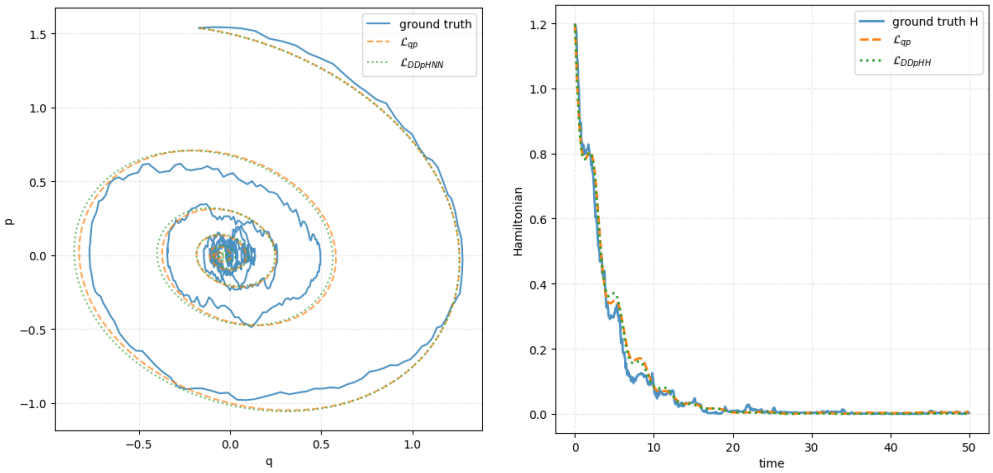


Fig. 3: Training-loss curves.

6.5.2 Comparing $\mathcal{L}_{VER}$ to a baseline Multi-Layer Perceptron (MLP)

The 3-DOF robotic arm

We consider a three-joint, three-degree-of-freedom (3-DOF) arm. For brevity, we only present plots for Joint 1; the results for Joints 2–3 are similar.



(a) Phase portrait (b) Energy evolution

Fig. 4: Phase portrait and corresponding energy evolution of the undamped mass-spring system.

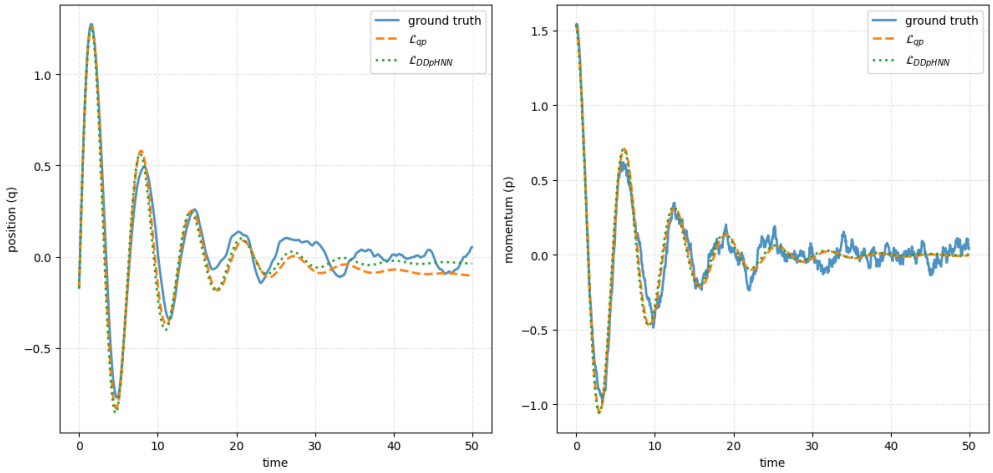


Fig. 5: Position and momentum.

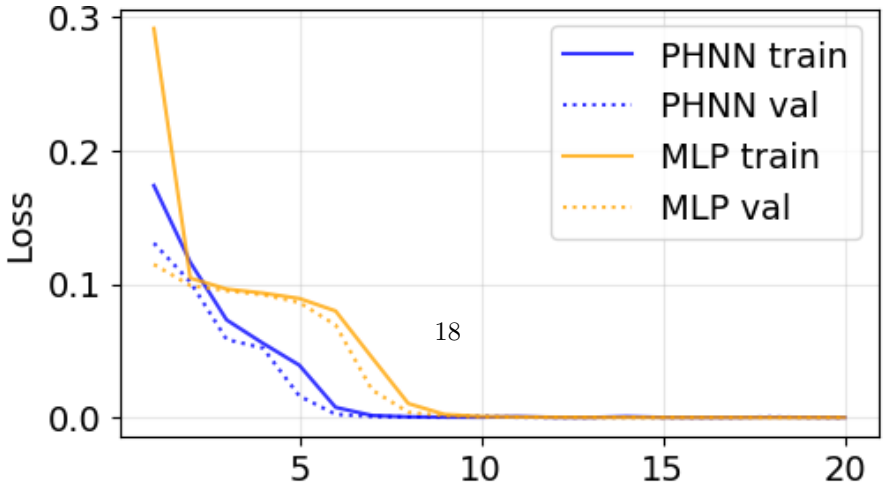
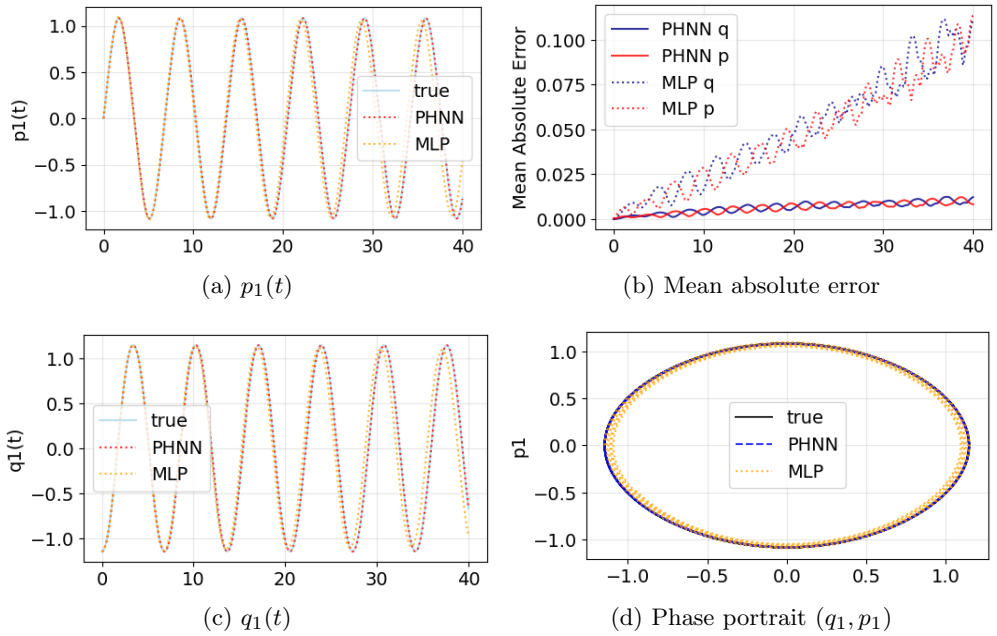


Fig. 6: Training-loss curves PHNN vs baseline model (MLP).

The PHNN converges faster and to a lower level than the MLP. The training and validation tracks are similar, indicating no overfitting.



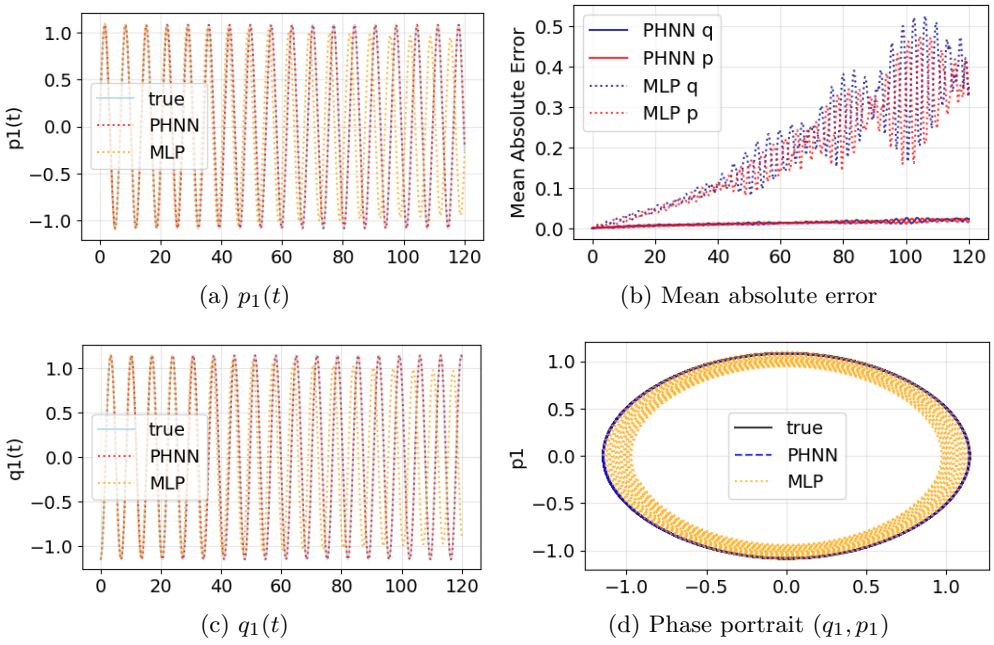
(a) $p_1(t)$ (b) Mean absolute error

(c) $q_1(t)$ (d) Phase portrait (q_1, p_1)

Fig. 7: PHNN vs MLP: time series, error, and phase portrait.

The PHNN trajectories are nearly indistinguishable from the ground truth. However, the MLP shows phase lag. The PHNN tracks amplitude and phase closely, while the MLP accumulates phase error. PHNN errors remain low and consistent, whereas MLP errors increase over time.

And now we do the same tests but for a longer time interval $[0, 120]$:



(a) $p_1(t)$ (b) Mean absolute error

(c) $q_1(t)$ (d) Phase portrait (q_1, p_1)

Fig. 8: PHNN vs MLP (extended): time series, error, and phase portrait.

The Duffing Oscillator

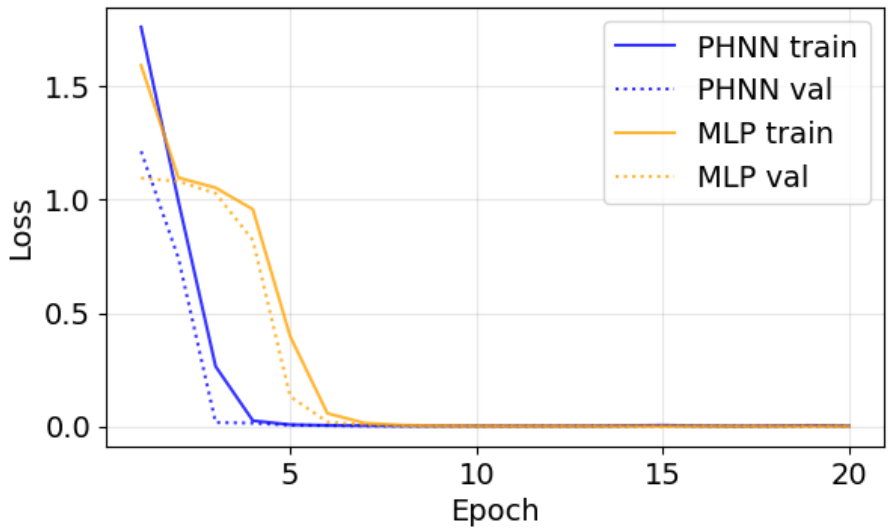


Fig. 9: Training-loss curves PHNN vs baseline model (MLP).

Both models drop quickly. PHNN converges faster and to a lower level. The training and validation curves track each other, indicating that there is no overfitting.

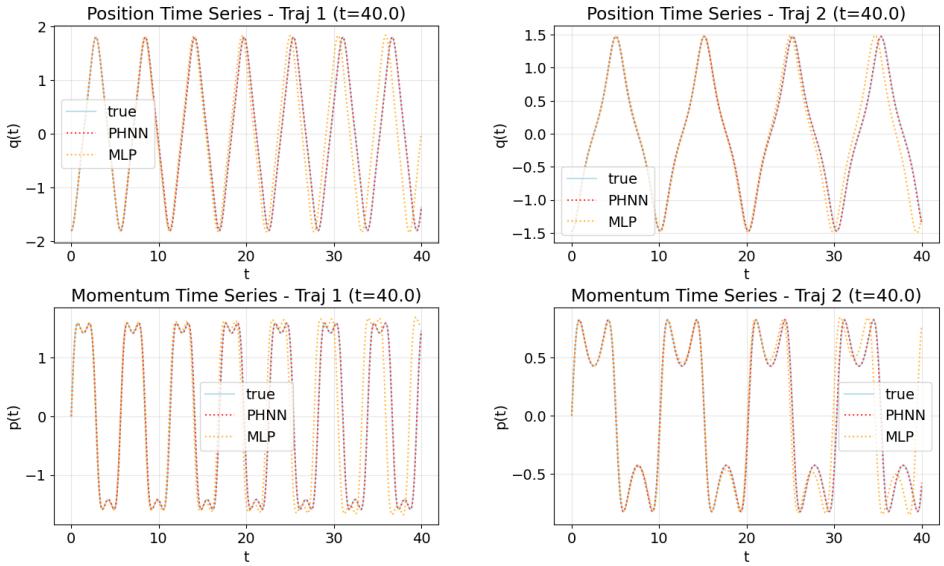


Fig. 10: Evolution of q and p for two different trajectories.

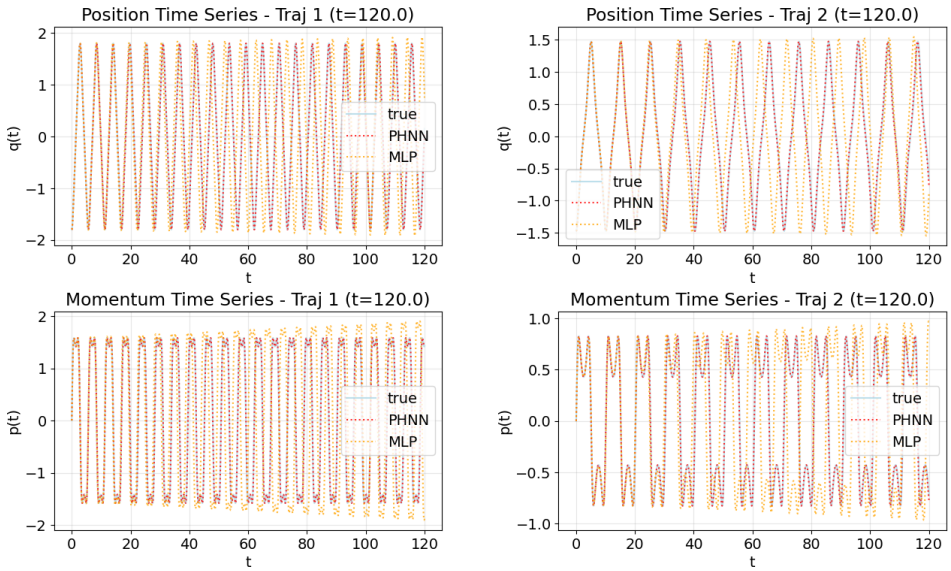


Fig. 11: Evolution of q and p for two different trajectories in a longer time interval $[0, 120]$.

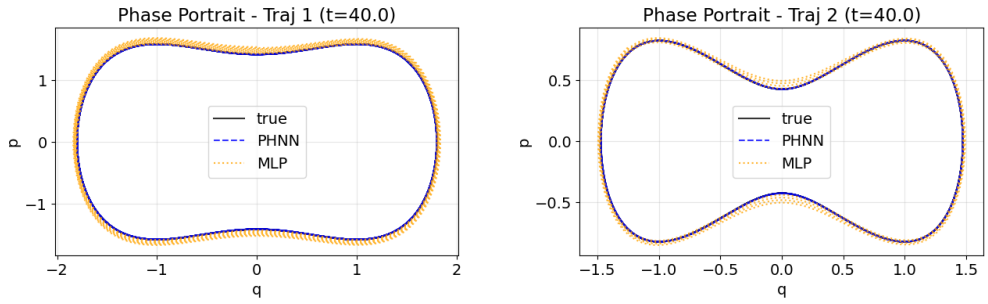


Fig. 12: Phase portrait for two different trajectories.

The PHNN is on top of the closed ground truth orbits. The MLP is close, but shows minor phase distortions near the turning points, hence demonstrating that both models have good short-horizon fidelity.

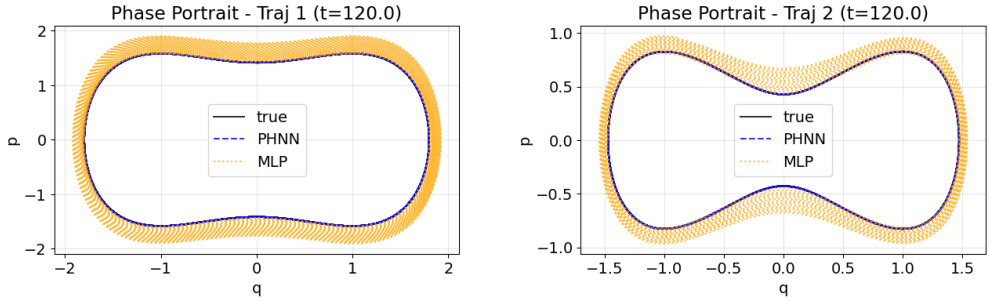


Fig. 13: Phase portrait for two different trajectories and a longer time interval $[0, 120]$

The PHNN aligns well. While the PHNN stays synchronized over the extended horizon, the MLP accumulates phase lag and amplitude drift.

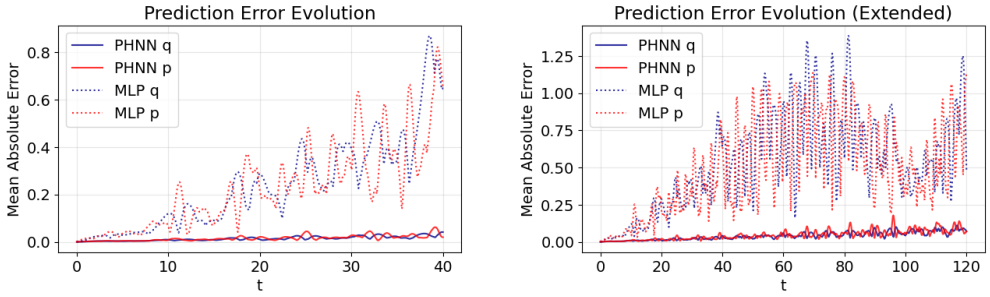


Fig. 14: Mean Absolute Error.

PHNN's q and p errors remain low and bounded over time, MLP's errors grow progressively, reflecting dephasing miscalibration during long rollouts.

7 Conclusions and Outlook

We revised the stochastic port-Hamiltonian framework so that the geometric structure remains a standard fiberwise Dirac structure, with stochasticity entering only through Stratonovich dynamics. This separation yields a pathwise Stratonovich energy identity, an Itô generator formula with an explicit diffusion correction term, and a passivity theory formulated in expectation rather than via deterministic pathwise bounds. Additionally, we demonstrated how weak passivity behaves under power-preserving interconnection and how the discrete discussion can be expressed in terms of finite-dimensional, maximally isotropic systems.

From a learning perspective, these results encourage the development of structure-preserving parameterizations and regularizers based on one-step consistency, corrected Itô energy drift, and generator-based passivity. Numerical experiments indicate that these constraints are compatible with accurate, long-term prediction of damped mass-spring, Duffing, and robotic benchmarks.

Future work includes Monte Carlo verification of the dissipativity inequality on learned trajectories, treatment of correlated-noise interconnections, and extensions to control synthesis and data assimilation settings, where both the Hamiltonian and diffusion structure must be identified from partial observations.

A Stochastic Neural Networks

Neurons are the basic computational units of the brain and form neural networks (NNs) through synaptic connections. Unlike deterministic NNs, real neurons introduce noise, resulting in probabilistic rather than fixed outputs. This stochastic behavior helps NNs avoid local minima during training, making them more robust to noisy or incomplete data. However, it increases the complexity of implementation and training. Moreover, the NNs inherent randomness also reduces overfitting by preventing the exact memorization of noisy training data.

Stochastic neural networks consist of interconnected neurons across layers. A key component is the stochastic neuron, which is often implemented using *magnetic tunnel junctions* (MTJs). An MTJ consists of two ferromagnets separated by a thin insulator and exhibits probabilistic switching behavior, see [18] for details. In neural network models, input spikes enter weighted synapses, as Figure 15 illustrates.

$$\Sigma := \text{Input} \times \text{Weights} = \text{Summed Output}$$

defines the weighted sum, after which the neuron fires based on a threshold or activation function.

Vreeken [15] introduced *Spiking Neural Networks* (SNNs), which closely mimic biological neurons. Neurons emit brief electrical spikes when sufficient input is accumulated. These spikes travel through axons and across synapses, comprising the axon terminal, synaptic gap, and dendrite, and convey information via spatial and temporal spike patterns. Yu et al. [17] proposed the *Simple and Effective Stochastic Neural*

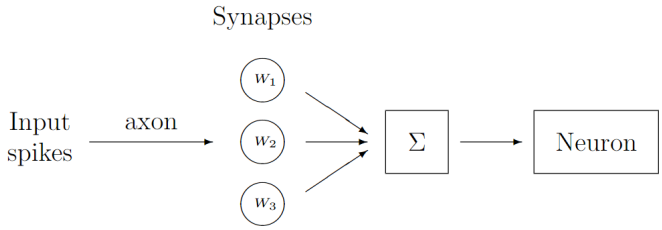


Fig. 15: Schematic representation of a basic artificial Neuron block.

Network (SE-SNN), which models activation uncertainty at each layer by predicting a Gaussian mean and variance and sampling during the forward pass. This approach improves robustness in pruning, adversarial defense, learning with label noise, and model calibration by leveraging an activation regularizer.

Rather than binary encoding, values can be represented by the probability of encountering ones in bit streams, which is central to stochastic computing (SC) [1]. SC uses random bit streams and standard digital logic for computation, allowing for simpler, fault-tolerant hardware. Functions such as sigmoid and tanh are implemented using linear finite-state machines, reducing cost at some precision's expense. As noted in [10], the noise in SC can also reduce overfitting and enhance inference accuracy. Both inputs and outputs are represented as probabilistic bit streams, merging stochastic methods with digital computation within a novel data-processing framework.

Acknowledgments

Matthias Ehrhardt acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 531152215 – CRC 1701.

References

[1] B.D. Brown and H.C. Card, Stochastic neural computation. I. Computational elements, IEEE Trans. Comput. 50(9) (2001), 891-905.

[2] F. Colonius and L. Grüne (eds.), Dynamics, Bifurcations and Control, Vol. 273, Springer Science & Business Media, 2002.

[3] F. Cordoni, L. Di Persio and R. Muradore, Stochastic port-Hamiltonian systems, J. Nonlin. Sci. 32(6) (2022), 91.

[4] S.A. Desai, M. Mattheakis, D. Sondak, P. Protopapas and S.J. Roberts, Port-Hamiltonian neural networks for learning explicit time-dependent dynamical systems, Phys. Rev. E 104(3) (2021), 034312.

[5] L. Di Persio, M. Ehrhardt and Y. Outaleb, Stochastic Port-Hamiltonian Neural Networks: Universal Approximation with Passivity Guarantees, arXiv:2603.10078, March 2026.

[6] M. Ehrhardt, Th. Kruse and A. Tordeux, The Collective Dynamics of a Stochastic Port-Hamiltonian Self-Driven Agent Model in One Dimension, ESAIM: Math. Model. Numer. Anal. 58 (2024), 515-544.

[7] C.W. Gardiner, Handbook of Stochastic Methods, Vol. 3. Berlin: Springer, 1985.

[8] S. Greydanus, M. Dzamba, and J. Yosinski, Hamiltonian neural networks. In: Advances in Neural Information Processing Systems, volume 32, 2019.

[9] P. Kotyczka and L. Lefevre, Discrete-time port-Hamiltonian systems: A definition based on symplectic integration, Systems & Control Letters 133 (2019), 104530.

[10] Y. Liu, S. Liu, Y. Wang, F. Lombardi and J. Han, A survey of stochastic computing neural networks for machine learning applications, IEEE Trans. Neural Networks Learning Syst. 32(7) (2020), 2809-2824.

[11] C. Neary, and U. Topcu, Compositional learning of dynamical system models using port-Hamiltonian neural networks, Learning for Dynamics and Control Conference, PMLR, (2023), 679-691.

[12] V. Talasila, J. Clemente-Gallardo and A.J. van der Schaft, Discrete port-Hamiltonian systems, Systems & Control Letters 55(6) (2006), 478-486.

[13] V. Talasila, J. Clemente-Gallardo and A.J. van der Schaft, Discrete port-Hamiltonian systems: mixed interconnections, Proceedings of the 44th IEEE Conference on Decision and Control, IEEE, 2005.

[14] A. van der Schaft and D. Jeltsema, Port-Hamiltonian systems theory: An introductory overview, Found. Trends Syst. Control 1(2-3) (2014), 173-378.

[15] J. Vreeken, Spiking neural networks, an introduction, (2003).

[16] C. Wang, S. Li, D. He, L. Wang, Is L^2 physics-informed loss always suitable for training an informed neural network?, Adv. Neural Inform. Process. Syst. 35 (2022): 8278-8290.

[17] T. Yu, Y. Yang, D. Li, T. Hospedales and T. Xiang, Simple and effective stochastic neural networks, Proc. AAAI Conf. Artificial Intelligence 35(4) (2021), 3252-3260.

[18] J.-G.J. Zhu and C. Park, Magnetic tunnel junctions, Materials Today 9(11) (2006), 36-45.

[19] M. David and F. Méhats, Symplectic learning for Hamiltonian Neural Networks, J. Comput. Phys. 494 (2023), 112495.

[20] F. Dietrich, A. Makeev, G. Kevrekidis, N. Evangelou, T. Bertalan, S. Reich, and I. G. Kevrekidis, Learning effective stochastic differential equations from microscopic simulations: linking stochastic numerics to deep learning, Chaos 33

(2023), 023121.

[21] M. Vaisband, V.vonBornhaupt, N. Schmid, I. Abulizi, J.Hasenauer, *etal.*, Loss formulations for assumption-free neural inference of SDE coefficient functions, npj Syst. Biol. Appl. 11 (2025), Article 22.