Bergische Universität Wuppertal

Fakultät für Mathematik und Naturwissenschaften

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

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

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

September 8, 2025

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,*], and Sofia Rizzotto[1]

[1] *Department of Computer Science – College of Mathematics, University of Verona, Italy.*
[2] *Chair of Applied and Computational Mathematics, University of Wuppertal, Gauß-strasse 20, 42119 Wuppertal, Germany.*
[3] *Doctoral School in Mathematics, University of Trento, Via Sommarive, 14 - 38123 Povo, Italy*

Abstract. This work introduces a new framework integrating port-Hamiltonian systems (PHS) and neural network architectures. This framework bridges the gap between deterministic and stochastic modeling of complex dynamical systems. We introduce new mathematical formulations and computational methods that expand the geometric structure of PHS to account for uncertainty, environmental noise, and random perturbations. Building on these advances, we introduce stochastic port-Hamiltonian neural networks (pHNNs), which facilitate the accurate learning and prediction of non-autonomous and interconnected stochastic systems.

Our proposed framework generalizes passivity concepts to the stochastic regime, ensuring stability while maintaining the system's energy-consistent structure. Extensive simulations, including those involving damped mass-spring systems, Duffing oscillators, and robotic control tasks, demonstrate the capability of pHNNs to capture complex dynamics with high fidelity, even under noise and uncertainty. This unified approach establishes a foundation for the robust, data-driven modeling and control of nonlinear stochastic systems.

1 Introduction

This research introduces a pioneering approach that combines *port-Hamiltonian systems* (PHS) with neural network architectures. The focus is on transitioning from de-

---

*Corresponding author.
Email:* luca.dipersio@univr.it (Luca Di Persio), ehrhardt@uni-wuppertal.de (Matthias Ehrhardt), youness.outaleb@unitn.it (Youness Outaleb), sofia.rizzotto@studenti.univr.it (Sofia Rizzotto)

terministic frameworks to stochastic models. Through an in-depth analysis of sophisticated mathematical formulations, we seek to enhance our comprehension of dynamical systems operating under uncertainty, encompassing a variety of intricate interactions and the potential for measurement errors.

PHS are defined as a synthesis of port modeling and geometric Hamiltonian dynamics, focusing on the Dirac structure. The Dirac structure generalizes traditional Poisson and pre-symplectic frameworks, enabling a nuanced representation of energetic topology within dynamical systems. PHS's stochastic adaptation incorporates uncertainties, such as inherent noise and environmental variabilities, directly into the system's ports. In this context, noise is not merely an external disturbance but an integral component of the system's behavior.

In our work, we offer a coordinate-free geometric formulation of deterministic PHS. This formulation employs generalized Poisson brackets and Hamiltonian functions to describe the underlying dynamics comprehensively. We seamlessly transition these theoretical constructs into local equations characterized by Jacobians and vector fields, providing a clear mathematical foundation. Furthermore, we extend this formulation to implicit systems with algebraic constraints by modeling energy storage and exchange using flow and effort variables and power ports.

The Dirac structure elucidates power-conserving interconnections within the system, which are essential for maintaining the principle of passivity. This property is crucial to stability in control applications. Building on this theoretical groundwork, we derive stochastic *port-Hamiltonian neural networks* (pHNNs) by incorporating noise within the port structure. This innovative framework is based on the methodology established by Cordoni, Di Persio, and Muradore [3], which adeptly integrates random perturbations to mitigate the impacts of measurement noise, parameter uncertainties, and environmental interactions, hence enriching the model's robustness.

The paper is organized as follows: Section 1 introduces key concepts and the various types of PHS, including discrete forms and their integration with neural networks. Basic concepts and definitions for PHS are introduced in Section 2. Section 3 extends PHS to include random elements for modeling systems with uncertainties. Section 4 generalizes the concept of *passivity* to stochastic systems by defining strong and weak passivity, as well as providing criteria for passivity in *stochastic port-Hamiltonian systems* (SPHS). Section 5 applies and develops stochastic PHS in different contexts. Section 6 presents simulation results of pHNNs on damped mass-spring and chaotic Duffing systems. Section 6 concludes with a summary and discusses how we adapted Colonius and Grüne's neural network–based controller design method for systems with stochastic dynamics [2].

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:

$$\begin{cases} \dot{x} & = X_H(x) + \sum_{i=1}^{m} u_i X_{H_{g_i}}(x), \\ y_i & = \{H, H_{g_i}\}. \end{cases} \tag{2.1}$$

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 (2.1) reads

$$\begin{cases} \dot{x} & = J(x)\partial_x H(x) + \sum_{i=1}^{m} u_i g_i(x), \\ y_i & = g_i^\top(x)\partial_x H. \end{cases} \tag{2.2}$$

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

$$\begin{cases} \dot{x} & = X_H^L(x) + \sum_{i=1}^{m} u_i H_{g_i}(x), \\ y_i & = [H, H_{g_i}], \end{cases} \tag{2.3}$$

and in local coordinates, this system (2.3) reads

$$\begin{cases} \dot{x} & = \big(J(x) - R(x)\big)\partial_x H(x) + \sum_{i=1}^{m} u_i g_i(x), \\ y_i & = g_i^\top(x)\partial_x H(x), \end{cases} \tag{2.4}$$

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, which can only transfer energy and cannot produce 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

$$v = -\dot{x} \quad \text{and} \quad v^* = \frac{\partial \mathcal{H}}{\partial x}(x), \tag{2.5}$$

implying the system is defined by

$$\left(-\dot{x}, \frac{\partial H}{\partial x}(x), f, e\right) \in \mathcal{D}(x). \tag{2.6}$$

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 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, only shifted between different forms and components. Ultimately, this relationship facilitates robust and 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

Until now, we have only considered port-Hamiltonian systems in continuous time. However, it is crucial to determine if their fundamental properties, such as energy conservation, are maintained during time discretization. As Viswanath, Clemente-Gallardo, and van der Schaft [12] demonstrated, Hamiltonian systems can be discretized to preserve energy-conserving behavior. Discrete port-Hamiltonian systems can be formed by discretizing continuous models or formulating them directly in discrete time. In the latter case, Poisson brackets are used to retain the critical structural properties necessary to preserve the Hamiltonian framework in the discrete domain: skew symmetry, bilinearity, and a modified Leibniz rule.

First, we must define the discrete Dirac structure. To do so, we denote a space of discrete vector fields by $\mathfrak{X}(A)$ and a space of discrete 1-forms by $\Lambda^1(A)$. Then, a *generalized Dirac structure* on an n-dimensional discrete manifold is an n-dimensional

linear subspace $\mathcal{D} \subset \mathfrak{X}(A) \times \Lambda^1(A)$ such that $\mathcal{D} = \mathcal{D}^\top$ with

$$\mathcal{D}^\top = \{(Y, \beta) \in \mathfrak{X}(A) \times \Lambda^1(A) \quad \text{where} \quad \langle \alpha, X \rangle + \langle \beta, Y \rangle = 0, \quad \forall\, (X, \alpha) \in \mathcal{D}\},$$

where $\langle \cdot, \cdot \rangle$ is the pairing between $\mathbb{F}^n$ and $\mathbb{F}^{n*}$. Consider the configuration of the effort-flow pairs shown in Figure 1. Let $\mathcal{F}_i$ denote the space of flow and $\mathcal{D}_i$ the space of effort of the Dirac structure for systems $i = A, B$. The *interconnection between the two Dirac structures* $\mathcal{D}_A$ and $\mathcal{D}_B$ can then be defined as

$$\mathcal{D}_A \circ \mathcal{D}_B := \{(f_1, e_1, f_2, e_2) \in \mathcal{F}_1 \times \mathcal{F}_1^* \times \mathcal{F}_2 \times \mathcal{F}_2^* \quad \text{such that}$$
$$\exists (f, e) \in \mathcal{F} \times \mathcal{F}^* \quad \text{with} \quad (f_1, e_1, f, e) \in \mathcal{D}_A \quad \text{and} \quad (-f, e, f_2, e_2) \in \mathcal{D}_B\}. \tag{2.7}$$

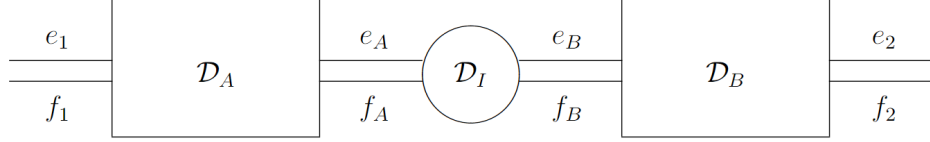Then $\mathcal{D}_A \circ \mathcal{D}_B$ is a Dirac structure.

Figure 1: Interconnection of two port-Hamiltonian systems.

Definition 2.1 (Implicit discrete port-Hamiltonian system [12]). *Let $H\colon \mathcal{Z} \to \mathbb{F}$ be a discrete Hamiltonian, $\mathcal{F}_P$ be the space of external flows f, $\mathcal{E}_P = \mathcal{F}_P^*$ be the space of external effort e, and $\mathcal{D}$ be the Dirac structure depending only on the coordinate z. Then, the implicit discrete port-Hamiltonian system is defined as*

$$\left(-\frac{\Delta z}{\Delta t}, f, \partial_z H(z), e\right) \in \mathcal{D}(z). \tag{2.8}$$

A system may consist of continuous and discrete port-Hamiltonian components encountered in various applications. To effectively manage these hybrid systems, Viswanath, Clemente-Gallardo, and van der Schaft [13] proposed a methodology that reinterprets energy-conserving interconnections at discrete sampling points as interconnections incorporating an external flow source. This transformation ensures energy conservation throughout sampling intervals, preserving the passive and port-Hamiltonian characteristics of the system.

This approach allows for integrating discrete computational models with continuous physical systems without compromising overall stability or passivity. However, it is noteworthy that this setup does not guarantee exact energy conservation because of the discrete sampling intervals.

To address the potential energy conservation issue, Kotyczka and Lefevre [9] introduced a discrete-time Dirac structure and a discrete-time port-Hamiltonian framework. Their methodology uses symplectic integration techniques, specifically collocation methods, to approximate the energy balance observed continuously. Their approach maintains the fidelity of the system's dynamics to the original continuous-time port-Hamiltonian representation by ensuring structural consistency over discrete time steps, allowing for the effective integration of numerical algorithms.

3   Stochastic port-Hamiltonian systems

In this section, we move to a stochastic setting. Random perturbations cause system trajectories to become stochastic processes. The deterministic duality pairing previously used to define Dirac structures must now be interpreted using *stochastic integrals*; moreover, it also necessitates a generalized definition of orthogonality that remains consistent with the energetic interpretation under stochastic dynamics.

Let $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \mathbb{R}_+}, \mathbb{P}\right)$ be a complete probability space. Denote the Stratonovich integral by δZ and the Itô integral by dZ along the semimartingale Z. As previously discussed, the stochastic port-Hamiltonian framework incorporates randomness by modeling each system component as a semimartingale. Using Stratonovich calculus aligns with the geometric structure of Dirac formulations. However, for analytical convenience and to leverage the probabilistic tools of stochastic analysis, this formulation can equivalently be expressed in terms of Itô calculus. Consider the system

$$\begin{cases} \delta X_t & = \left(J(X_t) - R(X_t) \right) \partial_x H(X_t) \, \delta Z_t + g(X_t) u \, \delta Z_t^g + \xi(X_t) \, \delta Z_t^N, \\ y_t & = g^\top(X_t) \, \partial_x H(X_t), \end{cases} \tag{3.1}$$

where $R(x) := (g^R(x))^\top \tilde{R}(x) g^R(x)$, W denotes a Brownian motion, and Z, Z^g, and Z^N are semimartingales. Then (3.1) describes the stochastic PHS in local coordinates, and the obtained framework can be further extended to scenarios in which noise enters the system by generalizing the concept of a Dirac structure. Noise can enter the system as a stochastic external field or as random perturbations affecting any of the system's connected ports.

Next, we introduce the *orthogonal complement* $\mathcal{D}^\perp$, consisting of all pairs $(\delta X_t, \sigma)$ that satisfy an energy-balance relation (expressed using *Stratonovich integrals*) concerning all elements of a given subbundle $\mathcal{D}$. The condition ensures no net energy is produced or lost by pairing these elements over time.

Definition 3.1 (Orthogonal complement, [3]). *Given a manifold $\mathcal{X}$, $I \subset \mathbb{R}_+$, a bundle $\mathcal{D} \subset T\mathcal{X} \oplus T^*\mathcal{X}$, a differential 1-form σ on $\mathcal{X}$ and an integral curve $X \colon I \to \mathcal{X}$ of a Stratonovich vector field δX_t, the orthogonal complement of $\mathcal{D}$ is*

$$\mathcal{D}^\perp = \left\{ (\delta X_t, \sigma) \subset T\mathcal{X} \oplus T^*\mathcal{X} \colon \int_0^t \langle \sigma, \delta \bar{X}_s \rangle + \int_0^t \langle \bar{\sigma}, \delta X_s \rangle = 0, \ \forall \, (\delta \bar{X}_t, \bar{\sigma}) \in \mathcal{D}, \ t \in I \right\} \tag{3.2}$$

From this, we define a *generalized stochastic Dirac structure* as a subbundle $\mathcal{D} \subset T\mathcal{X} \oplus T^*\mathcal{X}$ that equals its orthogonal complement: $\mathcal{D} = \mathcal{D}^\perp$. This condition ensures that $\mathcal{D}$ is both *isotropic* (energetically neutral) and *maximal* (it includes all such neutral pairs), thereby generalizing the classical concept of a Dirac structure to systems evolving under stochastic dynamics.

Definition 3.2 (Generalized Dirac structure, [3])**.** *With the same notation as above, we call generalized stochastic Dirac structure a smooth vector subbundle $\mathcal{D} \subset T\mathcal{X} \oplus T^*\mathcal{X}$ such that $\mathcal{D} = \mathcal{D}^\perp$.*

Remark 3.1. *We note that the above Definition 3.2 of generalized stochastic Dirac structures via Stratonovich pairing is not canonically anchored in geometric mechanics. Using energy-preserving pathwise integrals to define orthogonality lacks local (fiber-wise) meaning and bypasses the axiomatic isotropy and maximality central to Dirac theory.*
To resolve this issue, one can replace the integral criterion with a structure-preserving pairing induced by a stochastic symplectic form on the Itô tangent bundle. This pairing is defined as a bilinear form, denoted by $\omega_X\big((\delta X, \sigma), (\delta Y, \rho)\big) := \langle \rho, \delta X \rangle - \langle \sigma, \delta Y \rangle$, and construct D_x as the maximally isotropic subspace satisfying $\omega_X|_{D_x} \equiv 0$. This aligns with stochastic analogs of Courant algebroids.

We define an *implicit generalized stochastic port-Hamiltonian system* (IGSPHS), which naturally arises from extending Dirac structures to the stochastic setting. These generalized structures provide a geometric framework for capturing energy-conserving relations under stochastic perturbations. Classical PHSs use differential equations constrained by Dirac structures and driven by Hamiltonian functions representing stored energy. To model stochastic effects like noise or uncertainty, we introduce semimartingale perturbations and apply Stratonovich calculus, which aligns with the system's geometric nature. The IGSPHS ensures that the pair $\big(\delta X_t, \mathbf{d}H(X_t)\big)$ lies in the Dirac structure $\mathcal{D}(X_t)$ at each time, preserving energy consistency under randomness. Extending this formulation to include resistive and controlled ports allows for modeling dissipation and external interactions. Resistive elements are described by a subbundle $\mathcal{R}$, while controlled ports represent inputs and outputs.

Definition 3.3 (Implicit generalized stochastic PHS, [3])**.** *Let $H \colon \mathcal{X} \to \mathbb{R}$ be a Hamiltonian function, Z a semimartingale perturbing the system, then an implicit generalized stochastic port-Hamiltonian system on $\mathcal{X}$ is a 4-tuple $(\mathcal{X}, Z, \mathcal{D}, H)$ such that*

$$\big(\delta X_t, \mathbf{d}H(X_t)\big) \in \mathcal{D}(X_t) \quad \forall\, t \in I. \tag{3.3}$$

Including a resistive element and an external element control, then an implicit generalized port-Hamiltonian system with resistive structure $\mathcal{R}$ is a 5-tuple $(\mathcal{X}, \mathbf{Z}, \mathcal{F}, \mathcal{D}, H)$ such that

$$(-\delta X_t, \mathbf{d}H, \delta f_t^R, e_t^R, \delta f_t^C, e_t^C) \in \mathcal{D}(\mathcal{X})_\sqcup \quad with \quad (\delta f_t^R, e_t^R) \in \mathcal{R}(\mathcal{X})_t.$$

A significant advantage of the port-Hamiltonian framework is its ability to track energy flow within a dynamical system. When adapting the framework to a stochastic context, verifying that the fundamental energy-consistency property is maintained

is crucial. Proposition 3.1 establishes that *Interconnected Generalized Stochastic Port-Hamiltonian Systems* (IGSPHS) uphold a form of energy balance in both the pathwise and expectation senses.

The classical criterion of nonpositive energy dissipation through resistive ports is too strict in the presence of stochasticity. A more relaxed condition based on mean power balance ensures that energy dissipation occurs on average. This modification creates a robust, applicable framework for modeling stochastic energy exchange within the port-Hamiltonian paradigm.

Proposition 3.1 (Energy balance of IGSPHS, [3])**.** *Implicit port-Hamiltonian systems satisfy an energy conservation property that is*

$$H(X_t) - H(X_0) = \int_0^t \langle \boldsymbol{d}H, \delta X_s \rangle, \tag{3.4}$$

or, in short notation

$$\delta H(X_t) = \langle \boldsymbol{d}H, \delta X_t \rangle. \tag{3.5}$$

The energy balance is

$$H(X_t) - H(X_0) = \int_0^t \langle e_s^R, \delta f_s^R \rangle + \int_0^t \langle e_s^C, \delta f_s^C \rangle \leq \int_0^t \langle e_s^C, \delta f_s^C \rangle. \tag{3.6}$$

However, the condition $\int_0^t \langle e_s^R, \delta f_s^R \rangle \leq 0$ imposed on the resistive port is too strong since it is difficult for it to happen in practice, so the idea is to introduce the following weaker definition of the resistive relation $\mathcal{R}_W \subset \mathcal{F}_{Z_R} \times \mathcal{E}_R$:

$$\mathbb{E} \int_0^t \langle e_s^R, \delta f_s^R \rangle \leq 0. \tag{3.7}$$

Consequently, the mean power balance requires that the energy be conserved and dissipated in mean value, i.e. it reads

$$\mathbb{E}\big(H(X_t) - H(X_0)\big) = \mathbb{E} \int_0^t \langle e_s^R, \delta f_s^R \rangle + \mathbb{E} \int_0^t \langle e_s^C, \delta f_s^C \rangle \leq \mathbb{E} \int_0^t \langle e_s^C, \delta f_s^C \rangle. \tag{3.8}$$

It is possible to further generalize the Hamiltonian by introducing an external perturbation of the system, i.e. a new type of port called *noise port* perturbed by the semimartingale Z_N (see Figure 2). Thus, in this case, the *IGSPHS with resistive structure* is a 5-tuple $(\mathcal{X}, \mathbf{Z}, \mathcal{F}, \mathcal{D}, H)$ such that

$$\big(-\delta X_t, \boldsymbol{d}H, \delta f_t^R, e_t^R, \delta f_t^C, e_t^C, \delta f_t^N, e_t^N\big) \in \mathcal{D}(X_t), \tag{3.9}$$

and the weak energy balance is given by

$$\mathbb{E}H(X_t) - \mathbb{E}H(X_0) \leq \mathbb{E} \int_0^t \langle e_s^N, \delta f_s^N \rangle + \mathbb{E} \int_0^t \langle e_s^C, \delta f_s^C \rangle. \tag{3.10}$$

Furthermore, we can reformulate the SPHS in Itô form by defining the following alternative representation of the Dirac structure.
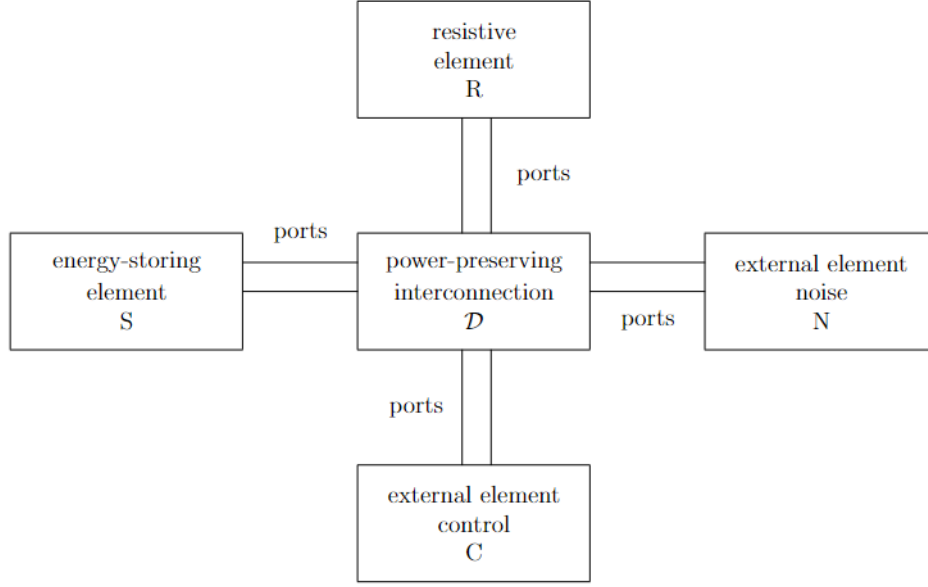
Figure 2: Schematic representation of a general implicit port-Hamiltonian system.

Definition 3.4 (Dirac structure, [3]). *Let $\mathcal{F} := \mathcal{F}_{Z_R} \times \mathcal{F}_{Z_C} \times \mathcal{F}_{Z_N}$ be the space of flows δf, $\mathcal{E} = \mathcal{F}^*$ be the dual space of efforts e, $G_\theta \colon \mathcal{F}_{Z_\theta} \to \mathcal{F}_{Z_\theta}$ be a function such that $\langle e_t^S, G_\theta \delta f_t^\theta \rangle = \langle G_\theta^* e_t^S, \delta f_t^\theta \rangle$ and J be a matrix with $J = -J^\top$. Then the Dirac structure $\mathcal{D}$ can be defined as*

$$\mathcal{D} := \big\{ (\delta f_t^S, \delta f_t^R, \delta f_t^C, \delta f_t^N, e_t^S, e_t^R, e_t^C, e_t^N) \in \mathcal{F} \times \mathcal{E} :$$
$$\delta f_t^S = -J e_t^S \delta Z_t - G_R \delta f_t^R - G_C \delta f_t^C - G_N \delta f_t^N, \tag{3.11}$$
$$e_t^R = G_R^* e_t^S, \quad e_t^C = G_C^* e_t^S, \quad e_t^N = G^* e_t^S \big\}.$$

Example 3.1. *Consider the special case*

$$\delta f_t^R = -\tilde{R} e_t^R \delta Z_t, \quad \delta f_t^N = \xi_t \delta Z_t^N, \quad \delta f_t^C = u_t \delta Z_t^C$$
$$\text{with} \quad \mathbb{E} \int_0^t \langle e_s^R, \tilde{R} e_s^R \delta Z_s \rangle - \mathbb{E} \int_0^t \langle e_s^N, f_s^N \delta Z_s^N \rangle \geq 0, \tag{3.12}$$

where the reason for the minus sign in front of $\tilde{R}$ is that we want it to be the incoming power regarding the interconnection (as in [14]).

Definition 3.5 (Stochastic input-output PHS with stochastic Dirac structure, [3]). *Using the same notation as above, if $\mathbf{Z} = (Z, Z^R, Z^C, Z^N)$ is a semimartingale and $H \colon \mathcal{X} \to \mathbb{R}$ is a Hamiltonian function, then the stochastic input-output port Hamiltonian system with stochastic Dirac structure is given by*

$$\begin{cases} \delta X_t & = -J dH(X_t)\, \delta Z_t + G_R \tilde{R} e_t^R\, \delta Z_t - G_C u_t\, \delta Z_t^C - G_N \xi_t\, \delta Z_t^N, \\ e_t^N & = G_N^* dH(X_t), \\ e_t^C & = G_C^* dH(X_t) \end{cases} \tag{3.13}$$

and by taking $\tilde{J} = -J$ and $e_t^R = G_R^* \mathbf{d}H(X_t)$, the system (3.13) becomes

$$
\begin{cases}
\delta X_t &= \left(\tilde{J} + G_R \tilde{R} G_R^*\right) \mathbf{d}H(X_t)\, \delta Z_t - G_C u_t\, \delta Z_t^C - G_N \xi_t\, \delta Z_t^N, \\
e_t^N &= G_N^* \mathbf{d}H(X_t), \\
e_t^C &= G_C^* \mathbf{d}H(X_t).
\end{cases}
\tag{3.14}
$$

Theorem 3.1 provides an explicit conversion from the Stratonovich formulation of an IGSPHS to its equivalent Itô representation. Under mild assumptions about the independence of the semimartingales driving the system, the resulting Itô formulation separates the contributions of the stochastic, resistive, and controlled components.

Theorem 3.1 (Conversion of IGSPHS, [3])**.** *If X is a solution of the equation* (3.14) *and* Z, Z^N, Z^C *are such that*

$$
\langle Z, Z^C\rangle_t = \langle Z, Z^N\rangle_t = \langle Z, Z^C\rangle_t = 0,
$$

where $\langle \cdot, \cdot \rangle_t$ *is the quadratic covariation at time t, then X can be equivalently rewritten in Itô terms as*

$$
\begin{aligned}
dX_t = {}& V^S(X_t)\, dZ_t + \mathcal{L}_{V^S} V^S(X_t)\, d\langle Z, Z\rangle_t \\
& - \sum_{i=1}^{n^N} V_i^N(X_t)\, dZ_t^N - \frac{1}{2}\sum_{i,j=1}^{n^N} \mathcal{L}_{C_j^N} V_i^N(X_t)\, d\langle Z^{N;i}, Z_t^{N;j}\rangle_t \\
& - \sum_{i=1}^{n^C} V_i^C(X_t) u_t^i\, dZ_t^{C;i} - \frac{1}{2}\sum_{i,j=1}^{n^C} \mathcal{L}_{V_j^C} V_i^C(X_t) u_t\, d\langle Z^{C;i}, Z^{C;j}\rangle_t,
\end{aligned}
\tag{3.15}
$$

where $\mathcal{L}$ *is the Lie derivative and* V^α, $\alpha = S, N, C$ *are defined as*

$$
\left(\tilde{J} + G_R \tilde{R} G_R^*\right) \mathbf{d}H = V^S, \quad G_N \xi_t = \sum_{i=1}^{n^N} V_i^M, \quad G_C = \sum_{i=1}^{n^C} V_i^C.
\tag{3.16}
$$

4   Passivity in Stochastic Systems

Extending passivity to stochastic systems is challenging because noise directly influences the energy dynamics of the system. In particular, the standard condition that the structure matrix R be symmetric and positive semidefinite is insufficient in the stochastic setting. The presence of a semimartingale Z can introduce energy influx, rendering the system non-dissipative. Moreover, we need additional constraints on the stochastic perturbations to preserve losslessness and passivity.

To analyze the behavior of observables in I-S-O stochastic PHS, it is essential to understand how smooth functions $\varphi \in C^\infty(\mathcal{X})$ evolve along the system trajectories. Proposition 4.1 shows that the evolution of such observables follows the same Lie bracket structure that defines the system dynamics. In this way, the geometric consistency of the Hamiltonian formulation is preserved under stochastic perturbations.

Proposition 4.1 (Evolution along system trajectories, [3]). *If X is the solution of an explicit I-S-O stochastic PHS with dissipation, i.e.*

$$\begin{cases} \delta X_t &= X_H^L(X_t)\,\delta Z_t + u X_{H_g}^L(X_t)\,\delta Z_t^g + X_{H_N}^L(X_t)\,\delta Z_t^N, \\ y_t &= [H, H_g]_L, \end{cases} \tag{4.1}$$

with

$$X_H^L(\cdot) := [\cdot, H]_L, \quad X_{H_g}^L(\cdot) := [\cdot, H_g]_L, \quad X_{H_N}^L(\cdot) := [\cdot, H_N]_L,$$

then for all $\varphi \in C^\infty(\mathcal{X})$ it holds

$$\begin{cases} \delta \varphi(X_t) &= [\varphi, H]_L(X_t)\,\delta Z_t + u[\varphi, H_g]_L(X_t)\,\delta Z_t^g + [\varphi, H_N]_L(X_t)\,\delta Z_t^N, \\ y_t &= [H, H_g]_L. \end{cases} \tag{4.2}$$

To assess the stability and energy behavior of stochastic PHSs, extending the classical concept of passivity to the stochastic setting is essential. We distinguish between two levels of passivity: strong passivity, which ensures energy bounds almost surely, and weak passivity, which guarantees these bounds in expectation, see Definition 4.1.

Definition 4.1 (Strong and weak passivity, [3]). *If $H \in C^\infty(\mathcal{X})$ is the total energy of the explicit I-S-O stochastic PHS with dissipation, then it is strongly passive if for all $t \geq 0$ holds*

$$H(X_t) \leq H(X_0) + \int_0^t u^\top(s) y(s)\,\delta Z_s^C, \tag{4.3}$$

or weakly passive if for all $t \geq 0$ it holds

$$\mathbb{E} H(X_t) \leq \mathbb{E} H(X_0) + \mathbb{E} \int_0^t u^\top(s) y(s)\,\delta Z_s^C. \tag{4.4}$$

Now, assuming $\xi = 0$, the energy conservation relation of the system

$$\begin{cases} \delta X_t &= \big(J(X_t) - R(X_t)\big)\partial_x H(X_t)\,\delta Z_t + g(X_t) u\,\delta Z_t^C + \xi(X_t)\,\delta Z_t^N, \\ e^C &= g^\top(X_t)\partial_x^H(X_t), \end{cases} \tag{4.5}$$

where $J = -J^\top$, is given by

$$\begin{aligned} H(X_t) - H(X_0) &= \int_0^t \langle \mathbf{d}H, \delta X_s \rangle = \int_0^t \langle e_s^R, \delta f_s^R \rangle + \int_0^t \langle e_s^C, \delta f_s^C \rangle \\ &= \int_0^t \langle \partial_x H(X_s), R(X_s) e^R\,\delta Z_s \rangle + \int_0^t \langle y, u\,\delta Z_s^C \rangle \\ &= \int_0^t \langle \partial_x H(X_s), -R(X_s)\partial_x H(X_s)\delta Z_s \rangle + \int_0^t y^\top u\,\delta Z_s^C \\ &= -\int_0^t \partial_x^\top H(X_s) R(X_s) \partial_x H(X_s)\delta Z_s + \int_0^t y^\top u\,\delta Z_s^C. \end{aligned} \tag{4.6}$$

Even if the matrix R is strictly positive, this alone does not guarantee strong passivity. In particular, we must also require the condition

$$\int_0^t \partial_x^\top H(X_s) R(X_s) \partial_x H(X_s) \, \delta Z_s \geq 0. \tag{4.7}$$

However, satisfying this condition in practice is often unrealistic. A more tractable alternative is to impose the weaker requirement

$$\mathbb{E} \int_0^t \partial_x^\top H(X_s) R(X_s) \partial_x H(X_s) \, \delta Z_s \geq 0, \tag{4.8}$$

which ensures *weak passivity*, cf. Def. 4.1. However, working directly with expectations of Stratonovich integrals can be technically challenging. To overcome this, it is natural to convert the expression in (4.4) into Itô form, by applying Theorem 3.1 to equation (4.4), to exploit related probabilistic tools then.

5   Application and Generalization

5.1   Interconnection of Multiple SPHS

An essential property of port-Hamiltonian systems (PHSs) is their interconnectivity, which allows complex systems to be viewed as compositions of simpler parts. The resulting interconnectivity regarding the components and how they are interconnected can be analyzed. In particular, through the composition of Dirac structures, the power-preserving interconnection of PHSs defines another PHS. The Hamiltonian of the interconnected PHS is the sum of the Hamiltonians of its components, and the energy-dissipation relation is the union of the energy-dissipation relations of the subsystems.

   The following Proposition 5.1 discusses the connection of multiple SPHS, defining a new system with interconnected Dirac structures and combined Hamiltonians.

Proposition 5.1. *Suppose we have N stochastic port-Hamiltonian systems with state space $\mathcal{X}_i$, Hamiltonian H_i, flow-effort space $\mathcal{F}_i \times \mathcal{E}_i$ and perturbation $\mathbf{Z}^i$ for $i = 1, \ldots, N$. Assuming that they are connected by $\mathcal{D}_I$ (see Figure 3), then their interconnection defines a stochastic port-Hamiltonian system with Dirac structure $\mathcal{D} \circ \mathcal{D}_I$ and Hamiltonian $H := \sum_{i=1}^N H_i$.*

5.2   Discrete Stochastic PHS

Consider the continuous stochastic port-Hamiltonian system

$$\begin{cases} dX_t &= \big((J - R)\partial_x H(X_t) + g(X_t)u_t\big) \, dt + \xi(X_t) \, \delta W_t, \\ y_t &= g^\top(X_t)\partial_x H(X_t), \\ z_t &= \xi^\top(X_t)\partial_x H(X_t), \end{cases} \tag{5.1}$$
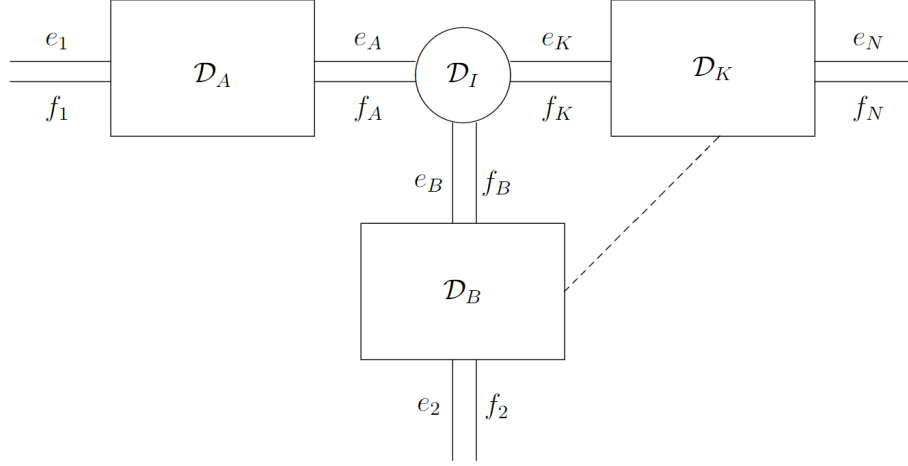
Figure 3: Interconnection of N implicit port-Hamiltonian systems.

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

$$\dot{X}(t_0^k + \tau h) = -f(t_0^k + \tau h) = -\sum_{j=1}^{s} f_j^k \, l_j(\tau), \tag{5.2}$$

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 ith 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 5.1 (Discrete stochastic port-Hamiltonian system, [9]). *A discrete stochastic port-Hamiltonian system can be written as*

$$\begin{cases} X(t_0^k + c_i h) &= x_0^k - h \sum_{i=1}^{s} a_{ij} f_j^k, \\ X(t_0^k + h) &= x_0^k - h \sum_{i=1}^{s} a_j f_i^k, \\ -a_{ij} f^k &= (J_j^k - R_j^k) \, a_{ij} e_j^k + a_{ij} g_j^k u_j^k + b_{ij} \xi_j^k \, \Delta W, \end{cases} \tag{5.3}$$

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

$$\mathbb{E}\left[\Delta H^k\right] \le h\,\mathbb{E}\left[(y^k)^\top u^k\right]. \tag{5.4}$$

5.3   Stochastic Motion Model of Agents

Ehrhardt, Kruse, and Tordeux present an application to a stochastic motion model of agents [6], who analyze the case where positions and velocities of agents are modeled in 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:

$$\begin{cases} dQ_n(t) &= \left(p_{n+1}(t) - p_n(t)\right) dt, \\ dp_n(t) &= \left(U'(Q_n(t)) - U'(Q_{n-q}(t))\right) dt \\ &\quad + \beta\left(p_{n+1}(t) - 2p_n(t) + p_{n-1}(t)\right) dt + \sigma\, dW_n(t), \end{cases} \tag{5.5}$$

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 nth 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 $Z(t) = \left(Q(t), p(t)\right)^\top \in \mathbb{R}^{2N}$, $t \in [0, +\infty)$, 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' behaviour under stochastic influences. In particular, the Hamiltonian is independent of Q and U, and its expectation could increase with 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

$$U(x) = \frac{(\alpha x)^2}{2} \quad x \in \mathbb{R},\ \alpha \in (0, \infty), \tag{5.6}$$

in which the process reads

$$dZ(t) = BZ(t)\, dt + G\, dW(t), \quad Z(0) = (Q_0, p_0)^\top, \tag{5.7}$$

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

6   Applications of Port-Hamiltonian Neural Networks

In this section, we examine the performance of *port-Hamiltonian neural networks* (pHNNs) on a mass-spring system, both with and without damping.

Here, pHNNs provide a robust framework for learning the dynamics of non-autonomous systems, which often involve time-dependent inputs and dissipative effects. These aspects typically pose challenges for conventional learning models.

As Desai et al. [4] have shown, pHNNs can accurately model such systems and capture complex behaviors. Notably, pHNNs can reconstruct Poincaré sections of chaotic systems, demonstrating their ability to learn the underlying structure from limited data, making pHNNs promising for applications involving nonlinear, forced, and damped systems, such as molecular dynamics, robotic control, and physical systems with unknown damping or input forces.
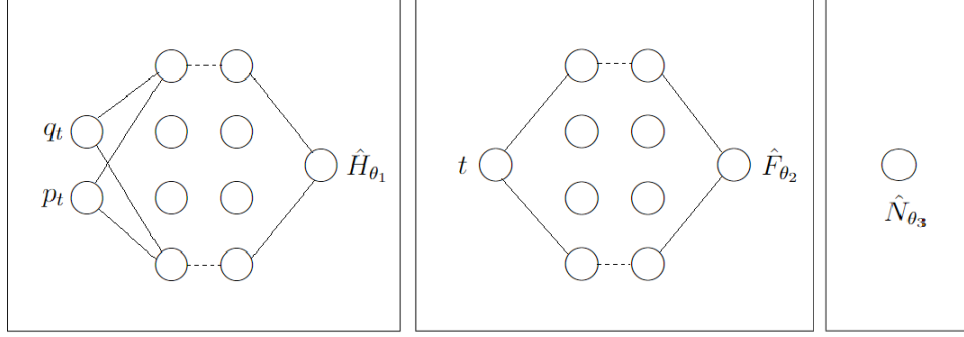

Figure 4: Schematic representation of a port-Hamiltonian neural network (pHNN).

As illustrated in Figure 4, the core idea is to leverage port-Hamiltonian theory to explicitly learn the force term F_{θ_2}, the damping term N_{θ_3}, and the Hamiltonian $\mathcal{H}_{\theta_1}$. This framework enables the prediction of the time derivatives:

$$\begin{bmatrix} \hat{\dot{q}}_t \\ \hat{\dot{p}}_t \end{bmatrix} = \begin{bmatrix} \frac{d\hat{H}_{\theta_1}}{dp_t} \\ -\frac{d\hat{H}_{\theta_1}}{dq_t} + \hat{N}_{\theta_3}\frac{d\hat{H}_{\theta_1}}{dp_t} + \hat{F}_{\theta_2} \end{bmatrix}. \tag{6.1}$$

6.1   First loss function

We can define a loss function using the observed time derivatives $[\dot{q}, \dot{p}]$ from the data, as described in [16]:

$$\mathcal{L}_{qp} = ||\hat{\dot{q}}_t - \dot{q}_t||_2^2 + ||\hat{\dot{p}}_t - \dot{p}_t||_2^2. \tag{6.2}$$

6.2   Second loss function

Recall that the following Itô SDE can model a stochastic port-Hamiltonian system

$$dX_t = f(X_t)\,dt + \Sigma(X_t)\,dW_t, \qquad X_t = [q_t^\top, p_t^\top]^\top,$$

with Brownian motion W_t. Over one sampling interval Δt the Euler–Maruyama increment $\Delta X_t = X_{t+\Delta t} - X_t$ satisfies $E[\Delta X_t] = f(X_t)\Delta t + O(\Delta t^2)$ and $\text{cov}[\Delta X_t] = \Sigma\Sigma^\top(X_t)\Delta t + O(\Delta t^2)$. Substituting neural predictions $\hat{f}_\theta$ and $\hat{\Sigma}_\theta$ for the unknown drift and diffusion yields the loss

$$\mathcal{L}_{\text{DDpHNN}} = \frac{1}{\Delta t}\|\Delta X_t - \hat{f}_\theta \Delta t\|_2^2 + \frac{1}{\Delta t}\|\Delta X_t \Delta X_t^\top - \hat{\Sigma}_\theta \hat{\Sigma}_\theta^\top \Delta t\|_F^2, \tag{6.3}$$

The drift-diffusion loss function (6.3) combines a drift term and a diffusion term:

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

We scale the second term of the drift-diffusion loss by λ to rebalance the training gradients and prevent the model from exaggerating the predicted noise magnitude, all while leaving drift learning unaffected. This results in well-calibrated diffusion without changing the network architecture

$$\mathcal{L}_{\text{DDpHNN}}^{(\lambda)} = \mathcal{L}_{\text{drift}} + \lambda\,\mathcal{L}_{\text{diff}}, \tag{6.4}$$

and to prevent the unbounded growth of Σ_θ, we add a ℓ_1-term ($\alpha << 1$) to control the diffusion amplitudes

$$\mathcal{L}_{\text{DDpHNN}}^{(\lambda,\alpha)} = \mathcal{L}_{\text{drift}} + \lambda\,\mathcal{L}_{\text{diff}} + \alpha\|\hat{\Sigma}_\theta\|_1. \tag{6.5}$$

Unlike standard HNN/pHNN training, which only accounts for deterministic drift and usually disregards stochasticity, (6.5) explicitly learns a state-dependent diffusion Σ_θ by matching the first two moments of Euler-Maruyama increments. This moment matching, together with the weighted and ℓ_1-regularized variant $\mathcal{L}_{\text{DDpHNN}}^{(\lambda,\alpha)}$, promotes calibrated noise magnitudes while leaving drift learning intact, hence providing one of the novel contributions of this paper. Typically, neural-SDE models learn drift and diffusion using either moment-matching of Euler-Maruyama increments or likelihood-based training. For example, [20] fit drift/diffusion via one-step EM maximum likelihood, while [21] proposes a hybrid objective that outperforms pure moment-matching. Our contribution is to introduce this approach into a port-Hamiltonian setting with a weighted moment-matching loss and an ℓ_1 penalty on Σ_θ, a method which, as far as we know, has not been reported before.

6.3   Third loss function

Let a minibatch $B = \{(x_i, f_i, h_i)\}_{i=1}^m$ be given, with the 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

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

We build a new loss function from the following three terms:

Vector field MSE

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

this serves to predict the Hamiltonian vector field $(\dot{q}, \dot{p}) = \left(\frac{\partial H}{\partial p}, -\frac{\partial H}{\partial q}\right)$ and minimize MSE to the ground-truth $(\dot{q}, \dot{p})$. This is the core loss used by Hamiltonian Neural Networks (HNN) [8] and follow-ups that supervise time derivatives rather than states.

Batch centered energy MSE

We regress the scalar energy $H(x)$, but we compare the 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 makes the regression immune to unknown offsets while teaching the network the shape and scale of the energy landscape.

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

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

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

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

and define

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

$\mathcal{L}_{\text{roll}}$ serves for long-horizon trajectory fidelity. The total loss is:

$$\mathcal{L}_{VER} = \mathcal{L}_{\text{vf}} + \lambda_{\text{E}}\,\mathcal{L}_{\text{E}} + \mathcal{L}_{\text{roll}}, \tag{6.6}$$

with $\lambda_{\mathrm{E}} > 0$ a tuning weight.

The VER objective ($\mathcal{L}_{VER}$) unifies three complementary signals-vector-field MSE, batch-centered energy regression, and a symplectic rollout penalty, into a single loss. To our knowledge this exact combination has not appeared before. This combination is interesting because it simultaneously constrains local derivatives, the global energy landscape, and long-horizon trajectory fidelity under a structure-preserving integrator. The VER loss combines three terms: a vector-field MSE that matches $(\dot{q}, \dot{p}) = J\nabla H_\theta$ to the observed derivatives, a batch-centered energy regression comparing $H_\theta - \bar{H}_\theta$ to $H_{\mathrm{true}} - \bar{h}$ to make the target invariant to additive constants while shaping the energy landscape, and a one-period symplectic rollout penalty that enforces long-horizon fidelity. The first component follows the HNN paradigm of supervising time derivatives, as proposed by H [8], and the third uses "symplectic in the loss" training with structure-preserving schemes [19]. To our knowledge, unifying all three in a single objective is novel, as it jointly constrains local dynamics, global energy structure, and multi-step stability.

We will now test the three objectives. The resulting learning curves, phase portraits, energy traces, state trajectories, and error statistics are presented in the following figures and discussed below.

6.4   Tests and Results

6.4.1   Comparing $\mathcal{L}_{qp}$ and $\mathcal{L}_{DDpHNN}$:

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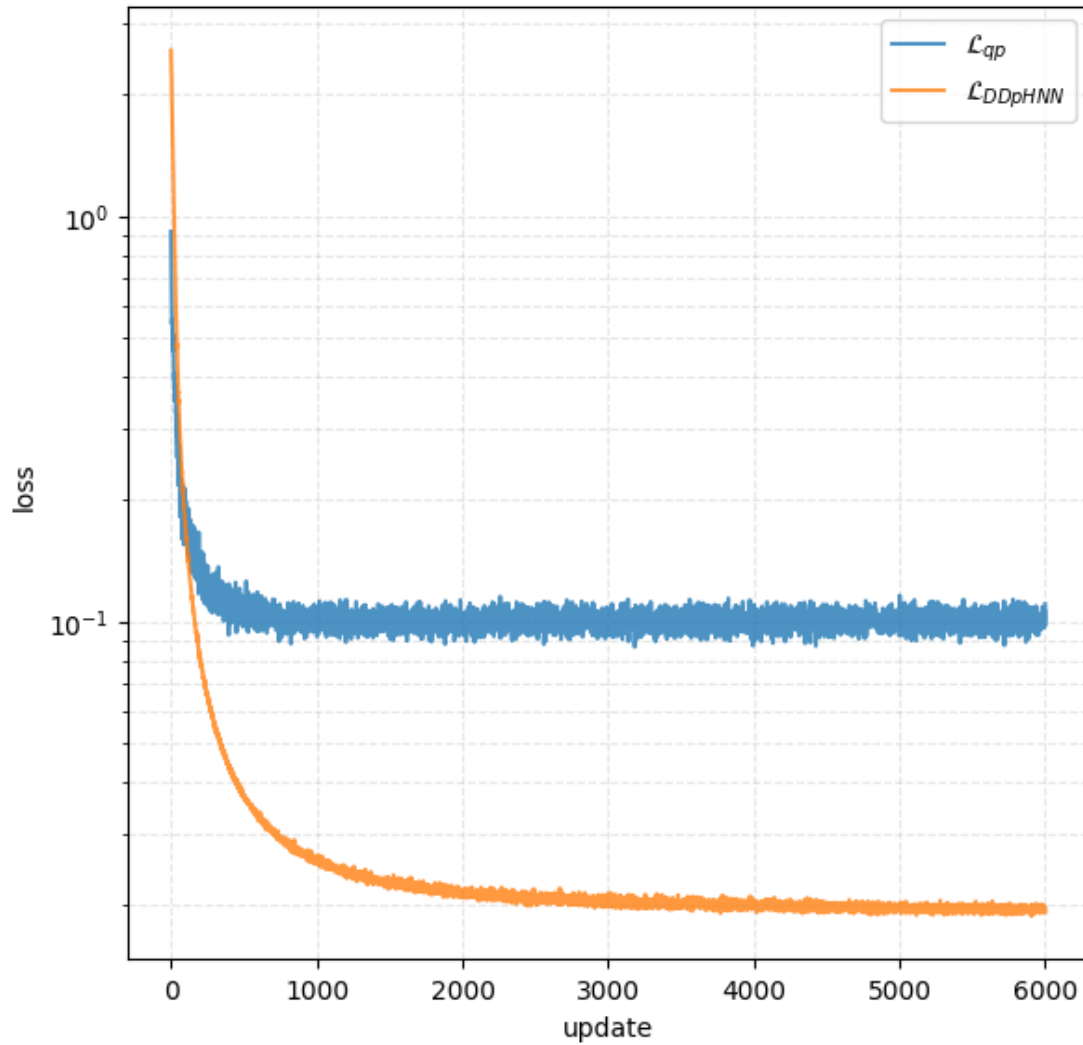
Training loss

Figure 5: Training-loss curves.

Figure 5 confirms stable convergence for both objectives. The loss $\mathcal{L}_{\text{DDpHNN}}$ (orange) decays one order of magnitude below the deterministic loss $\mathcal{L}_{\text{qp}}$ (blue) and plateaus without signs of overfitting. This shows that the additional diffusion term is well behaved.
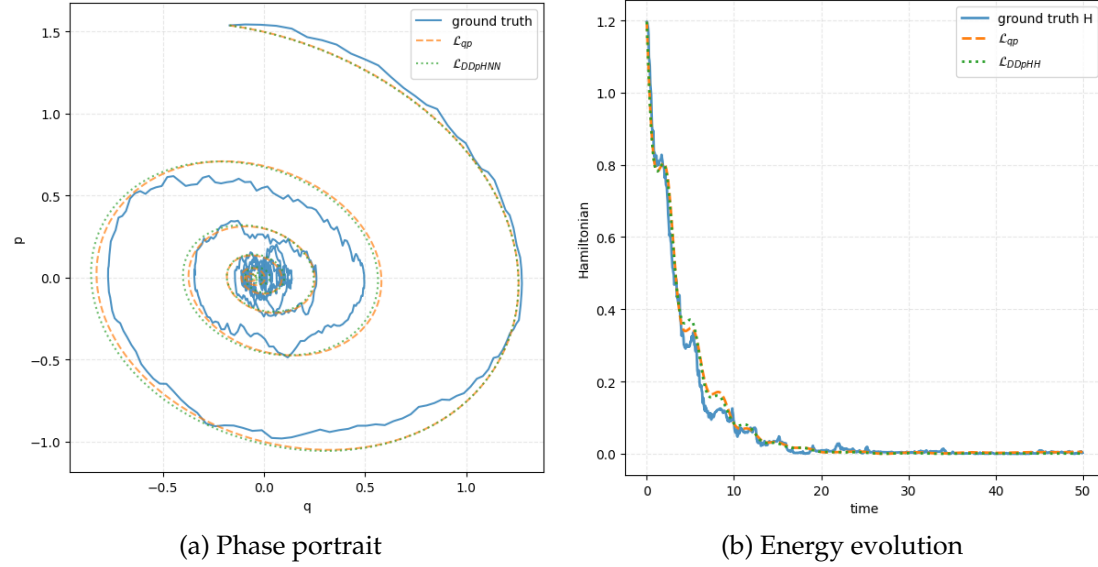
Phase portrait and energy evolution

(a) Phase portrait   (b) Energy evolution

Figure 6: Phase portrait and corresponding energy evolution of the undamped mass-spring system.

As shown in Figure 6a, both models closely follow the true spiral trajectory for the entire time window. Figure 6b tracks the Hamiltonian $H(q, p)$. Both networks exhibit the dissipative trend.

Position and momentum
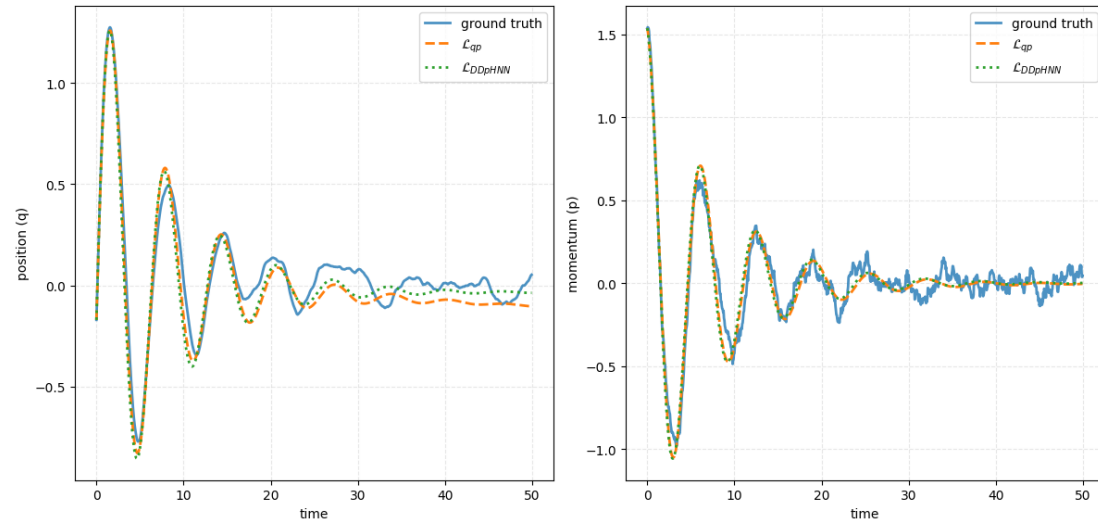
Figure 7: Position and momentum.

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

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

The 3-DOF robotic arm

We consider a 3-DOF arm (three joints). For brevity, we present Joint 1 plots, results for Joints 2–3 are similar.
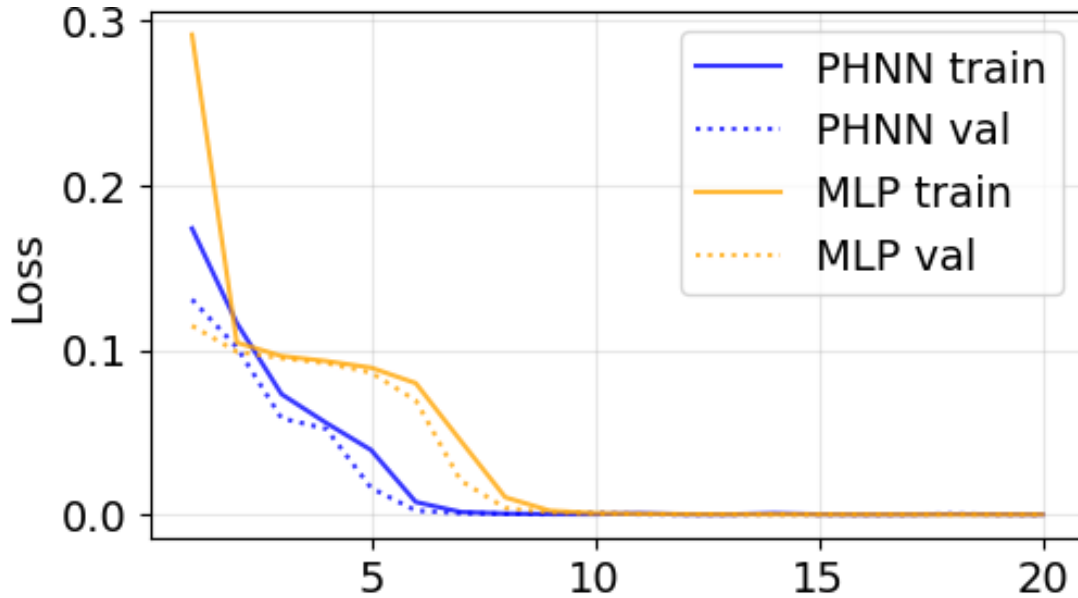

Figure 8: Training-loss curves PHNN vs baseline model (MLP).

PHNN converges faster and lower than the MLP, training and validation track closely which indicate no overfitting.

(a) $p_1(t)$

(b) Mean absolute error

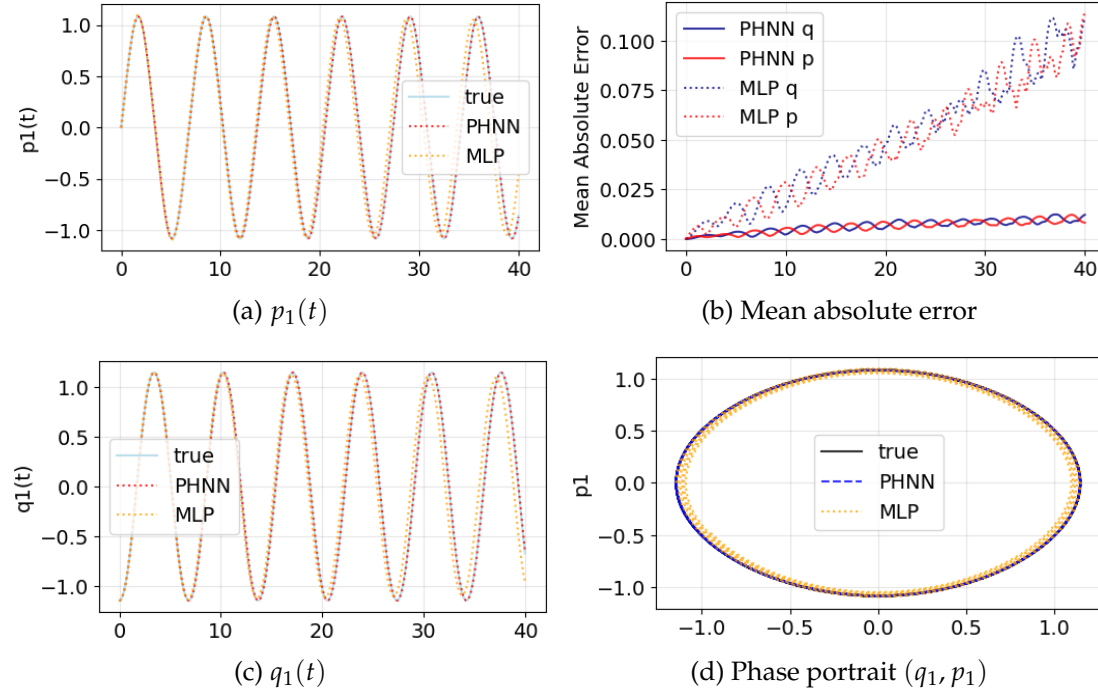(c) $q_1(t)$

(d) Phase portrait (q_1, p_1)

Figure 9: 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.

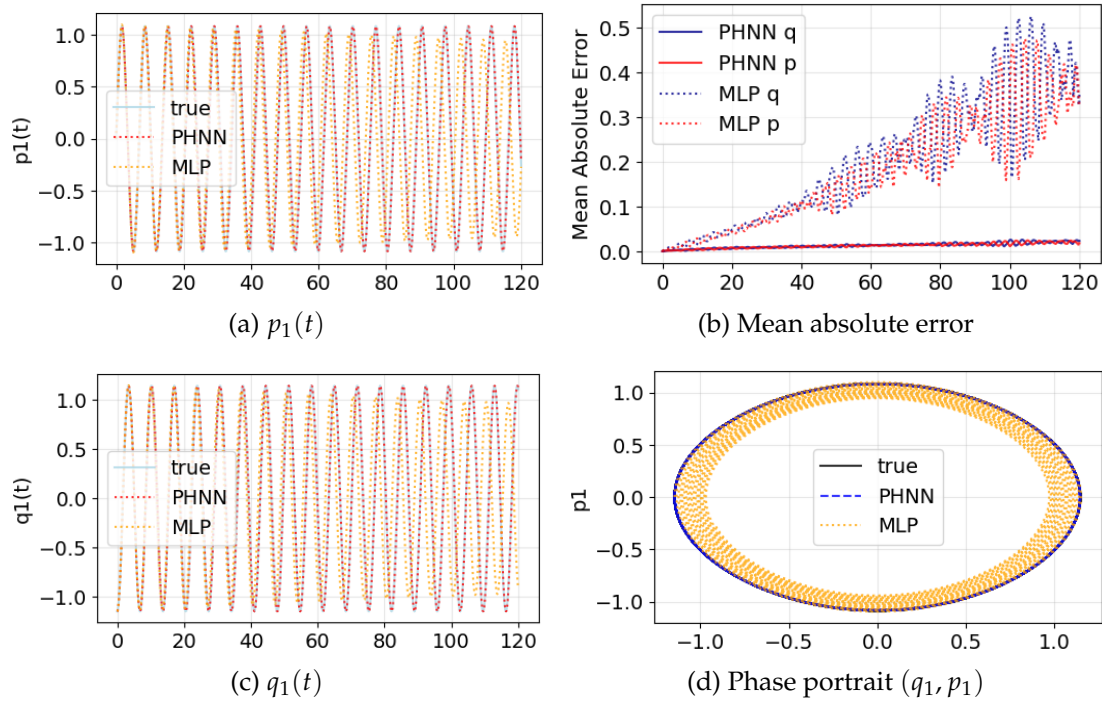And now we do the same tests but for a longer time interval $[0, 120]$:

(a) $p_1(t)$

(b) Mean absolute error

(c) $q_1(t)$

(d) Phase portrait (q_1, p_1)

Figure 10: PHNN vs MLP (extended): time series, error, and phase portrait.
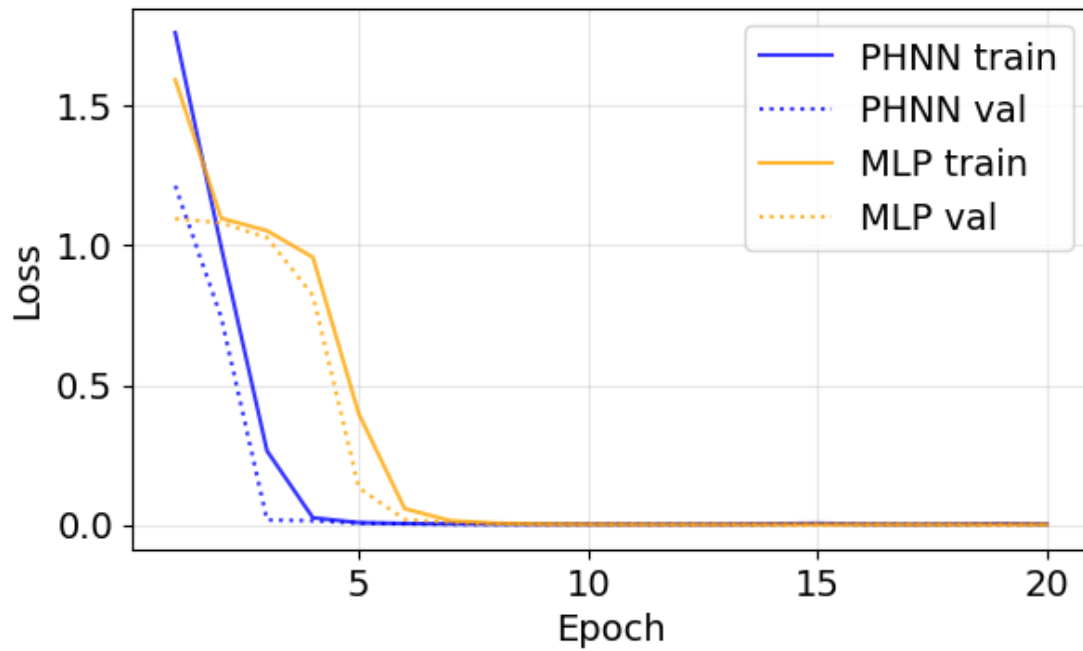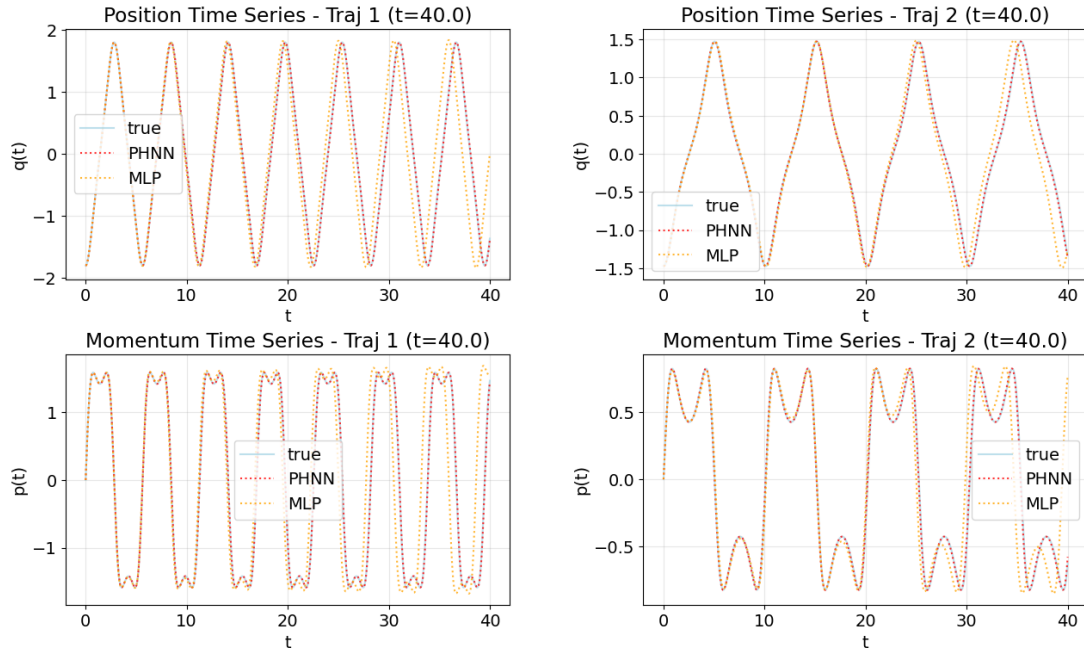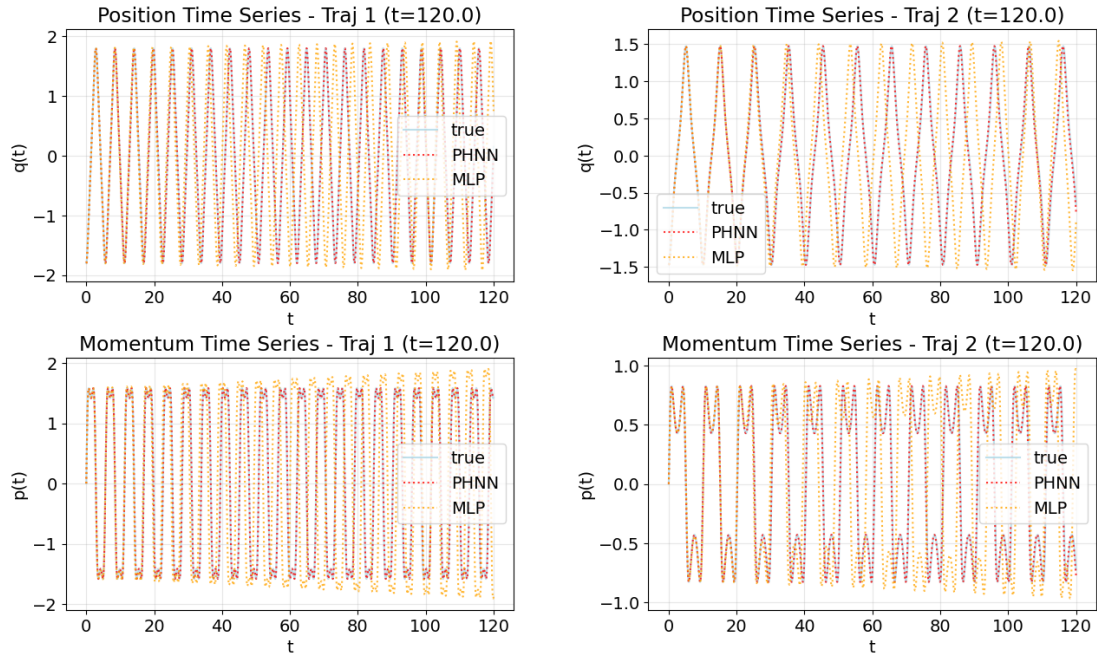
The Duffing Oscillator

Figure 11: 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.

Figure 12: Evolution of q and p for two different trajectories.

Figure 13: Evolution of q and p for two different trajectories in a longer time interval $[0, 120]$.
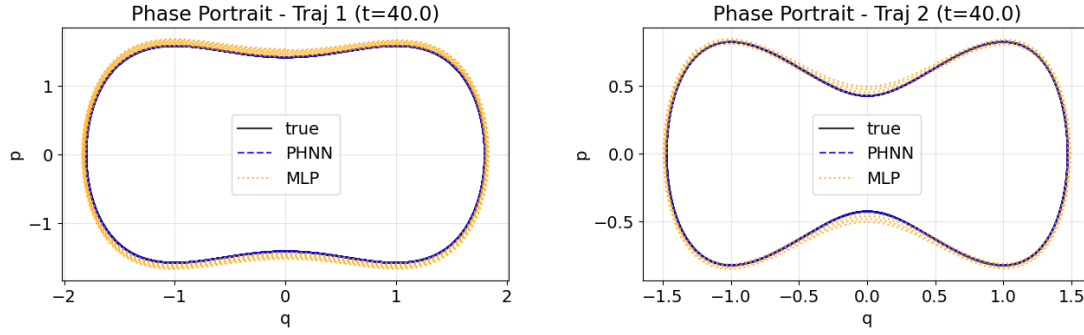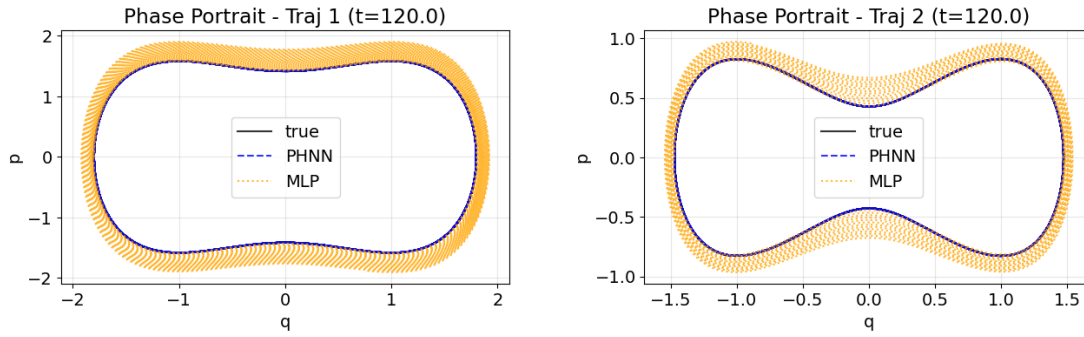
Figure 14: Phase portrait for two different trajectories.

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.

Figure 15: 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.
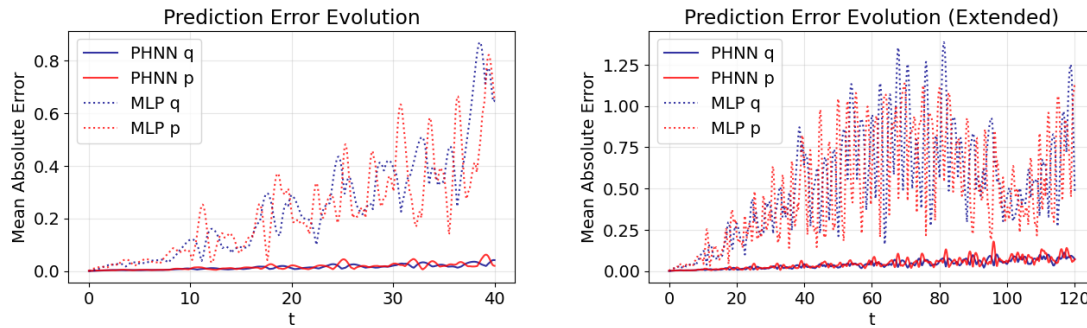

Figure 16: Mean Absolute Error.

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

The method of Colonius and Grüne [2] uses a neural network to estimate a gradient field for controller design, removing the necessity to solve complicated partial differential equations. To extend this framework to stochastic port-Hamiltonian systems (SPHS), however, we must incorporate stochastic elements into the system model and the control synthesis.

First, we defined a stochastic version of the port-Hamiltonian system using stochastic differential equations (SDEs), such as:

$$dx = \big(J(x) - R(x) \big) \nabla H(x)\, dt + G(x)u\, dt + \Sigma(x)\, dW,$$

Here, $\Sigma(x)$ is a matrix that captures the noise intensity, and dW represents a Wiener process. Second, the control strategy must be adapted to handle stochasticity. A modified control law can be written as follows

$$u = -K(x) + \mu(x)$$

Here, $K(x)$ is derived from the deterministic system, and $\mu(x)$ compensates for stochastic effects. A suitable Lyapunov function is constructed and analyzed using Itô's calculus to ensure system stability. The objective is to demonstrate that the expected value of the Lyapunov function decreases over time, guaranteeing stability under stochastic dynamics. Evolutionary optimization strategies should be adapted for this stochastic setting by redefining the fitness function to reflect expected performance across a range of noise conditions. The ball-and-beam system will be extended to include stochastic disturbances as a practical demonstration. Simulations will illustrate the effects of noise and validate the proposed stochastic control approach.

An important direction for future research is to extend our conversion framework beyond the assumption of pairwise uncorrelated stochastic drivers, as required by Theorem 3.1. This assumption excludes a wide class of practical systems with inherently coupled control and noise inputs. Future work could address this limitation and enable a more general and realistic modeling of coupled stochastic systems by incorporating non-orthogonal stochastic drivers using the full cross-variation tensor, $\langle Z^i, Z^j \rangle$. This approach would require applying Itô's Lemma in manifold coordinates equipped with a nontrivial Levi-Civita connection.

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 intro-

duce 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 17 illustrates.
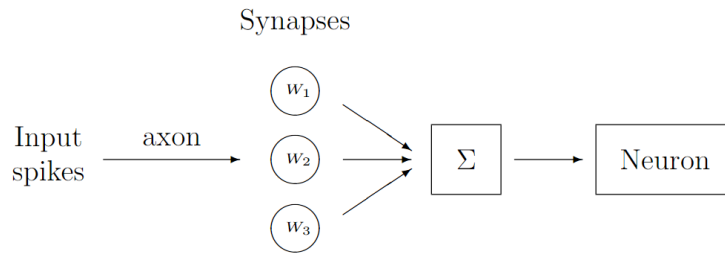

Figure 17: Schematic representation of a basic artificial Neuron block.

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

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 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 with the help of 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 via 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 in a novel data processing framework.

Acknowledgments

Some Remarks and Suggestions

On the Validity of Proposition 3.1 and the Role of Coisotropy

In Proposition 3.1, the stochastic energy balance for an implicit generalized stochastic port-Hamiltonian system is formulated via Stratonovich integration:

$$H(X_t) - H(X_0) = \int_0^t \langle \nabla H(X_s), \delta X_s \rangle, \tag{A.1}$$

while this pathwise identity is formally correct in the framework of Stratonovich calculus, it implicitly assumes that the stochastic perturbations are compatible with the geometry of the Hamiltonian, in the sense that they do not induce any net drift in the energy evolution. To rigorously ensure the validity of the stochastic energy balance (A.1), we must verify that the stochastic dynamics satisfy a *coisotropy condition*:

$$\nabla H(x)^\top \Sigma(x) = 0 \quad \text{for all } x \in \mathcal{X}, \tag{A.2}$$

where Σ is the diffusion matrix of the Itô representation of the system, hence asserting that the stochastic perturbations act tangentially to the level sets of the Hamiltonian. Therefore, any diffusion-driven motion does not change the energy of the system. Geometrically, the diffusion vector fields must lie in the coisotropic subspace orthogonal to ∇H, thereby preserving the energy structure of the Hamiltonian dynamics. Without this assumption (A.2), the stochastic integral in (A.1) may possess a non-zero drift. Therefore the identity may not hold in expectation or even in distribution. This issue becomes particularly relevant when attempting to apply probabilistic tools, as in the case of Dynkin's formula, which relies (also) on Itô calculus, hence requiring knowledge of the drift and diffusion terms of the process.

To overcome these limitations of the pathwise formulation and to ensure compatibility with probabilistic analysis, we reformulate the energy balance in the weak (expectation) sense using the infinitesimal generator of the underlying Itô diffusion.

Let us consider a stochastic port-Hamiltonian system (SPHS) evolving under the Itô SDE:

$$dX_t = f(X_t)\,dt + \Sigma(X_t)\,dW_t, \tag{A.3}$$

where $X_t \in \mathbb{R}^n$ is the state vector, $f\colon \mathbb{R}^n \to \mathbb{R}^n$ is the drift vector field, $\Sigma\colon \mathbb{R}^n \to \mathbb{R}^{n \times m}$ is the diffusion matrix, and W_t is an m-dimensional Brownian motion adapted to the filtration $(\mathcal{F}_t)_{t \geq 0}$. In the port-Hamiltonian formulation, the drift vector f has the structure:

$$f(x) = \big(J(x) - R(x)\big)\nabla H(x) + G(x)u(t), \tag{A.4}$$

where (as before):

- $H\colon \mathbb{R}^n \to \mathbb{R}$ is the Hamiltonian, i.e., the energy function,

- $J(x)$ is a skew-symmetric structure matrix representing the interconnection (satisfying $J(x)^\top = -J(x)$),

- $R(x)$ is a symmetric positive semi-definite matrix modeling dissipation,

- $G(x)$ is the control port matrix,

- $u(t)$ is the control input.

Accordingly, we compute the infinitesimal generator $\mathcal{L}$ of the diffusion process X_t applied to the observable H. Since $H \in C^2$, the generator reads:

$$\mathcal{L}H(x) = \nabla H(x)^\top f(x) + \frac{1}{2} \operatorname{tr}\left(\Sigma(x)\Sigma(x)^\top \nabla^2 H(x)\right), \tag{A.5}$$

and, under the coisotropy condition (A.2), the first term $\nabla H^\top \Sigma$ vanishes identically. Consequently, the diffusion term in (A.5) does not contribute to the drift of $H(X_t)$ in the Itô expansion. Therefore, we have:

$$\mathcal{L}H(x) = \nabla H(x)^\top \left(J(x) - R(x)\right)\nabla H(x) + \nabla H(x)^\top G(x)u(t), \tag{A.6}$$

and via the skew-symmetry of J, we observe that:

$$\nabla H(x)^\top J(x)\nabla H(x) = 0,$$

and hence:

$$\mathcal{L}H(x) = -\nabla H(x)^\top R(x)\nabla H(x) + y(x)^\top u(t), \tag{A.7}$$

where $y(x) = G(x)^\top \nabla H(x)$ is the output port variable conjugate to the input $u(t)$.

Now, applying Dynkin's formula to the process $H(X_t)$, we obtain the *weak energy balance*:

$$\frac{d}{dt}\mathbb{E}\left[H(X_t)\right] = \mathbb{E}\left[\mathcal{L}H(X_t)\right] = -\mathbb{E}\left[\nabla H(X_t)^\top R(X_t)\nabla H(X_t)\right] + \mathbb{E}\left[y_t^\top u_t\right]. \tag{A.8}$$

Moreover, if $R(x) \succeq 0$ for all x, then the first term in (A.8) is non-positive, and we obtain the inequality:

$$\frac{d}{dt}\mathbb{E}\left[H(X_t)\right] \leq \mathbb{E}\left[y_t^\top u_t\right], \tag{A.9}$$

which expresses the *weak passivity* of the system in the sense of expected energy. Let us underline that Equation (A.9) provides a rigorous foundation for the expected energy dissipation of SPHSs and constitutes a more robust and mathematically sound alternative to the pathwise identity (A.1): it shows that energy dissipation in expectation is guaranteed under two verifiable *structural assumptions*:

1. The damping matrix $R(x)$ is positive semi-definite;

2. The coisotropy condition (A.2): $\nabla H^\top \Sigma = 0$ holds globally.

These hypotheses ensure that the stochastic perturbations do not inject energy into the system on average, and that the deterministic dissipation encoded in $R(x)$ governs the energy decay.

Remark A.1. *We also remark that the passivity inequality* (A.9) *extends naturally to more general settings, including time-varying or control-dependent Hamiltonians, provided suitable regularity and compatibility conditions are satisfied. Furthermore, this framework enables a natural definition of a* storage function *and the use of stochastic Lyapunov techniques to assess stability in expectation.*

Accordingly to what above, we could think about addressing the missing verification of coisotropy (A.2) in Proposition 3.1, providing the following amendments to the paper:

- Introduce the coisotropy condition (A.2) explicitly as a structural assumption required for the validity of the energy balance (A.9).

- Add a new proposition (as above) that derives the energy balance in the weak sense using the Itô generator and Dynkin's formula.

- Optionally, retain Proposition 3.1 as a formal identity under the Stratonovich calculus, but clearly delimit its domain of validity and contrast it with the expectation-based result.

What is above preserves the original Stratonovich formulation's geometric intuition, enhances the paper's analytic rigor, and bridges the gap with stochastic control theory. Moreover, we align the generalized stochastic Dirac structures introduced earlier in the manuscript, while ensuring consistency across the geometric and probabilistic formulations.

Point (5) - Passivity in the Stochastic Regime (detailed). Let us start by reminding the reader that strong passivity is declared through the pathwise inequality

$$H(X_t) \leq H(X_0) + \int_0^t u^\top(s)\, y(s)\, \delta Z_s^C, \qquad\qquad (\mathrm{P_s})$$

where the storage is the Hamiltonian H and the supply rate is $u^\top y$. For the stochastic PHS

$$\delta X_t = \big(\mathbf{J} - \mathbf{R}\big)\nabla H(X_t)\, \delta Z_t + \mathbf{G}(X_t)u_t\, \delta Z_t^C + \Sigma(X_t)\, \delta W_t,$$

the ensuing energy balance (4.6) contains the Stratonovich line integral $\int_0^t \nabla H^\top \mathbf{R}\, \nabla H\, \delta Z_s$. Imposing its *pathwise* non-negativity (condition (4.7)) is generically impossible, even with $\mathbf{R} \geq 0$, because the signed measure δZ_s may reverse sign along a single trajectory. Moreover, concerning the *Mean-Square Passivity* (MSP), replace $(\mathrm{P_s})$ by the *mean-square* dissipativity inequality

$$\mathbb{E}\Big[H(X_t)\Big] - \mathbb{E}\Big[H(X_0)\Big] \leq \mathbb{E}\left[\int_0^t u^\top(s)\, y(s)\, ds\right] \qquad\qquad (\mathrm{P_{ms}})$$

together with the *coisotropy constraint*

$$\nabla H(x)^\top \Sigma(x) = 0, \qquad \forall x \in X, \tag{C_Σ}$$

which forces the diffusion to act tangentially to the energy level sets. And we can state a result of the following type

Theorem A.1 (MSP Sufficiency). *Assume $H \in C^2$, $\mathbf{R}(x) \succeq 0$, and (C_Σ). Let $\mathcal{L}$ be the Itô generator $\mathcal{L}f = \frac{1}{2}\operatorname{tr}(\Sigma\Sigma^\top \nabla^2 f) + \left[(\mathbf{J} - \mathbf{R})\nabla H + \mathbf{G}u\right]^\top \nabla f$. Then*

$$\mathcal{L}H = -\nabla H^\top \mathbf{R}\, \nabla H \leq u^\top y,$$

and consequently inequality (P_{ms}) holds for all $t \geq 0$.

    Sketch of a possible proof Apply Dynkin's formula to $H(X_t)$ and exploit (C_Σ) to cancel the diffusion term; since $\mathbf{J}$ is skew, $\nabla H^\top \mathbf{J} \nabla H \equiv 0$, and the remainder yields the stated bound.

    Storage Functions & Martingale Structure. Under MSP, $V(t) := H(X_t) - \int_0^t u^\top y\, ds$ is a supermartingale: $\mathbb{E}[V(t)] \leq \mathbb{E}[V(0)]$, hence H acts as a Lyapunov function *in expectation*; stability follows by Doob's supermartingale convergence theorem. Concerning what we can define as the problem of designing guidelines for Data-Driven models, we have [I know what follows is very sketchy, but we deepen the arguments, in case we decide to harm ourselves :)]

1. *Positive damping.* Parameterize $\mathbf{R}_\theta = \mathbf{L}_\theta \mathbf{L}_\theta^\top$ (Cholesky or softplus), guaranteeing $\mathbf{R}_\theta \succeq 0$ during training.

2. *Noise orthogonality.* Constrain Σ_θ via a projection $\Sigma_\theta(x) := \left(\mathbf{I} - \frac{\nabla H_\theta \nabla H_\theta^\top}{\|\nabla H_\theta\|^2}\right)\widetilde{\Sigma}_\theta(x)$, which enforces ($C_\Sigma$).

3. *Loss augmentation.* Add $\lambda \|(\nabla H_\theta)^\top \Sigma_\theta\|_2^2$ to the training loss and fit $\lambda \gg 0$ adaptively.

    Interconnection. For two MSP systems $(\Sigma_1, \mathbf{R}_1, H_1)$ and $(\Sigma_2, \mathbf{R}_2, H_2)$ interconnected via power ports, the composite storage $H := H_1 + H_2$ remains MSP, because the supermartingale property is additive and the supply rates cancel internally. This generalizes the deterministic passivity theorem to the stochastic domain.

    Final considerations
Let us emphasize that strong (pathwise) passivity cannot be upheld for generic diffusion-driven port-Hamiltonian systems. Consider any semimartingale Z_t with unbounded variation, such as Brownian motion or a Lévy process. The Stratonovich integral, which appears in the energy balance, is given by $I = \int_0^t \nabla H^\top R\, \nabla H\, \delta Z_s$. Even when $\mathbf{R} \succeq 0$, the integrand is nonnegative, while the differential δZ_s changes sign infinitely

often on almost every trajectory. Consequently, the integral itself oscillates and may be positive or negative on arbitrarily short intervals. A pointwise inequality of the form

$$H(X_t) \leq H(X_0) + \int_0^t u^\top(s)\, y(s)\, \delta Z_s^C$$

is therefore violated with probability one, unless Z_t has bounded variation. This justifies abandoning strong passivity in favor of a weaker, yet attainable, notion. Moreover, *mean-square passivity* (MSP) replaces the pathwise energy constraint with an expectation inequality

$$\mathbb{E}\big[H(X_t)\big] - \mathbb{E}\big[H(X_0)\big] \leq \mathbb{E}\Big[\int_0^t u^\top(s)\, y(s)\, ds\Big]. \tag{MSP}$$

This inequality (MSP) states that the *expected* stored energy never exceeds the supplied work. Indeed, by writing the Itô dynamics in the form $dX_t = f(X_t, u_t)\, dt + \Sigma(X_t)\, dW_t$ with $f = (\mathbf{J} - \mathbf{R})\nabla H + \mathbf{G}u$ and imposing the orthogonality constraint $\nabla H^\top \Sigma \equiv 0$, the Itô generator becomes $\mathcal{L}H = -\nabla H^\top \mathbf{R}\,\nabla H + u^\top y$, and since $\mathbf{R}$ is positive semidefinite, $\mathcal{L}H \leq u^\top y$, and Dynkin's formula yields (MSP) automatically. Hence the requirement *boils down* to two constructive conditions: positivity of the damping matrix $\mathbf{R}$ and covariance-orthogonality of the diffusion Σ concerning the Hamiltonian gradient. Such conditions can be encoded directly in data-driven (neural) surrogates. A common strategy is to parameterize the damping through a Cholesky factor, $\mathbf{R}_\theta = L_\theta L_\theta^\top$, guaranteeing $\mathbf{R}_\theta \succeq 0$ for all network parameters. Orthogonality of the diffusion can be enforced by projecting a free network output $\widetilde{\Sigma}_\theta$ onto the tangent space of the energy level sets,

$$\Sigma_\theta(x) = \Big(\mathbf{I} - \frac{\nabla H_\theta \nabla H_\theta^\top}{\|\nabla H_\theta\|_2^2 + \varepsilon}\Big)\, \widetilde{\Sigma}_\theta(x),$$

where ε is a small constant preventing numerical blow-up when ∇H_θ is near zero. In practice, one augments the training loss with large penalties on both the violation of the orthogonality condition and the discrepancy $\mathcal{L}_\theta H_\theta - u^\top y_\theta$. After training, long-horizon roll-outs can be used to verify empirically that the map $t \mapsto \mathbb{E}[H_\theta(X_t)] - \mathbb{E}\big[\int_0^t u^\top y_\theta\, ds\big]$ is non-increasing, confirming that MSP holds in simulation.

Unlike the unattainable strong passivity constraint, MSP is achievable for diffusion-driven systems, retains the intuitive *energy cannot be created on average* interpretation, remains closed under interconnection of subsystems (the sum of two MSP storage functions is again MSP), a perfect (well done) puzzle with statistical learning where expectations are estimated empirically. Summing up, unless we want to keep chasing unicorns in the Stratonovich jungle, we might as well accept that MSP isn't just a convenient compromise; it's the only passivity notion that won't collapse under the weight of actual stochasticity.

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, manuscript in preparation.

[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, Proceedings of the AAAI Conference on 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.