2022 IEEE 61st Conference on Decision and Control (CDC)
December 6-9, 2022. Cancún, Mexico

Gaussian Process Port-Hamiltonian Systems: Bayesian Learning with Physics Prior

Thomas Beckers[1], Jacob Seidman[2], Paris Perdikaris[3], George J. Pappas[1]

2022 IEEE 61st Conference on Decision and Control (CDC) | 978-1-6654-6761-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/CDC51059.2022.9992733

***Abstract*— Data-driven approaches achieve remarkable results for the modeling of complex dynamics based on collected data. However, these models often neglect basic physical principles which determine the behavior of any real-world system. This omission is unfavorable in two ways: The models are not as data-efficient as they could be by incorporating physical prior knowledge, and the model itself might not be physically correct. We propose Gaussian Process Port-Hamiltonian systems (GP-PHS) as a physics-informed Bayesian learning approach with uncertainty quantification. The Bayesian nature of GP-PHS uses collected data to form a distribution over all possible Hamiltonians instead of a single point estimate. Due to the underlying physics model, a GP-PHS generates passive systems with respect to designated inputs and outputs. Further, the proposed approach preserves the compositional nature of Port-Hamiltonian systems.**

I. Introduction

The modeling and identification of dynamical systems is a crucial task in a broad range of domains such as physics, engineering, applied mathematics and medicine [1]–[3]. System identification has a long history in control as most control strategies are derived based on a model of the plant. Whereas classical models of physical systems are typically based on first principles, there is a recent shift to more data-driven modeling of complex dynamics to capture more details with reduced engineering effort. However, this paradigm shift poses new questions regarding the efficiency, interpretability, and physical correctness of these models [4]. By physical correctness we mean that the learned model respects physical principles such as conservation of energy and passivity. Including these physical principles in a data-driven approach is beneficial in several ways: The models 1) are more meaningful as they respect the postulates of physics, 2) come with increased interpretability, and 3) can be more data-efficient as satisfaction of physical axioms results in a meaningful inductive bias of the model [5].

A large class of physical systems can be described by Hamiltonian mechanics, see [6]. The main idea is that the total energy in the system is described by the Hamiltonian, a smooth scalar function of the system's generalized coordinates. However, Hamiltonian systems are restricted to dynamics without external inputs and dissipative elements, which is rarely the case in real-world scenarios.

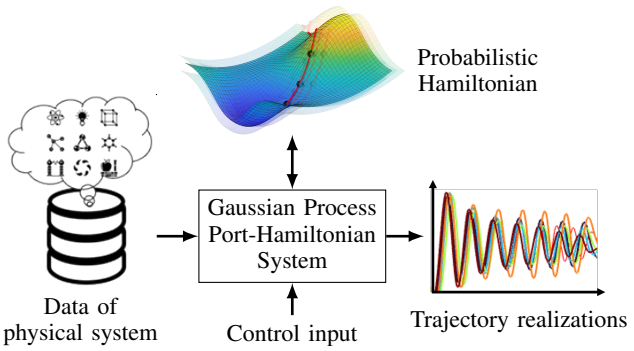


Fig. 1. Gaussian Process Port-Hamiltonian systems allow to learn a probabilistic Hamiltonian based on collected data. The Port-Hamiltonian structure enables to include control inputs as well as dissipation. All sample trajectories of the GP-PHS model are physically correct in terms of the evolution of the total energy in the system.

A composable framework to overcome these restrictions has been provided by Port-Hamiltonian systems (PHS), see [7]. PHS leverage network approaches from electrical engineering and constitute a cornerstone of mathematical systems theory. While most of the analysis of physical systems has been performed within the Lagrangian and Hamiltonian framework, the network point of view is attractive for the modeling and simulation of complex physical systems with many connected components. PHS consider three kinds of ideal components: 1) energy-storing elements, 2) energy-dissipating (resistive) elements, and 3) energy-routing elements. It has been shown that PHS are able to model a range of multi-domain complex physical systems involving combinations of electrical, mechanical, electromechanical, chemical, hydrodynamical, thermodynamical, and quantum components, see [8]. Therefore, PHS form an expressive class of systems that are able to model many real world physical scenarios of interest.

Results for the data-driven learning of Hamiltonian systems with neural networks have been established in [9], [10], and extended to PHS, see [11], [12]. However, standard neural network architectures can have difficulty generalizing from sparse data and have no uncertainty quantification of the outputs. In contrast, Bayesian models such as Gaussian processes (GP) can generalize well even for small training datasets and have a built-in notion of uncertainty quantification [13]. By using the Bayesian approach to learn a posterior

This work was supported by the AFOSR under grant FA9550-19-1-0265 (Assured Autonomy in Contested Environments)

[1] are with the Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104, USA, tbeckers@seas.upenn.edu, pappasg@seas.upenn.edu

[2] is with the graduate program in Applied Mathematics and Computational Science, University of Pennsylvania, Philadelphia, PA 19104, USA, seidj@sas.upenn.edu

[3] is with the Department of Mechanical Engineering and Applied Mechanics, University of Pennsylvania, Philadelphia, PA 19104, USA, pgp@seas.upenn.edu

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 14:44:34 UTC from IEEE Xplore. Restrictions apply.

distribution of functions, we are able to both obtain point predictions of outputs as well as a quantified uncertainty. GPs have been recently used for learning generating functions for Hamiltonian systems [14] or the dynamics of other structured systems [15]–[17], but they have not been applied to the class of Port-Hamiltonian systems.

Contribution: In this paper, we embed the powerful PHS representation into a Bayesian framework and propose Gaussian Process Port-Hamiltonian systems (GP-PHS) for learning physical systems based on state measurements. For this purpose, we introduce a PHS kernel for the GP that generates a distribution over PHS dynamics. The integration of physical principles allows the GP-PHS to be data-efficient and the resulting models are physically correct in the sense of the evolution of the system's energy. In contrast to existing PHS identifications methods, the probabilistic model includes all possible realizations of the PHS under the GP prior with respect to a finite set of data. We give a proof that the interconnection properties of PHS are preserved in GP-PHS which is a beneficial characteristic for system composition and control.

The remainder of the paper is structured as follows. We start with the introduction of PHS and GPs in Section II, followed by proposing GP-PHS in Section III. Finally, a simulation highlights the benefits of GP-PHS in Section IV.

II. Problem Formulation

In this section we briefly describe the class of Port-Hamiltonian systems, revisit GP regression, and introduce the problem setting.

A. Port-Hamiltonian Systems

Composing Hamiltonian systems with input/output ports leads to a Port-Hamiltonian system (PHS) which is a dynamical system with ports that specify the interactions of its components. The dynamics of a PHS Σ are described by[1].

$$\Sigma(J,R,H,G)=\begin{cases}\dot{\boldsymbol{x}}=[J(\boldsymbol{x})-R(\boldsymbol{x})]\nabla_{\boldsymbol{x}}H(\boldsymbol{x})+G(\boldsymbol{x})\boldsymbol{u}\\ \boldsymbol{y}=G(\boldsymbol{x})^\top\nabla_{\boldsymbol{x}}H(\boldsymbol{x}),\end{cases} \tag{1}$$

with the state $\boldsymbol{x}(t) \in \mathbb{R}^n$ (also called energy variable) at time $t \in \mathbb{R}_{\geq 0}$, the total energy represented by a smooth function $H\colon \mathbb{R}^n \to \mathbb{R}$ called the Hamiltonian, and the I/O ports $\boldsymbol{u}, \boldsymbol{y} \in \mathbb{R}^m$.

The matrix $J\colon \mathbb{R}^n \to \mathbb{R}^{n\times n}$ is skew-symmetric and specifies the interconnection structure and the matrix $R\colon \mathbb{R}^n \to \mathbb{R}^{n\times n}, R = R^\top \succeq 0$ specifies the dissipation in the system. The contact with the environment is defined by the matrix $G\colon \mathbb{R}^n \to \mathbb{R}^{n\times m}$. The structure of the interconnection matrix J is typically derived from kinematic constraints in mechanical systems, Kirchhoff's laws, power transformers, gyrators, etc. Loosely speaking, the interconnection of the elements in the PHS is defined by J, whereas the Hamiltonian H characterizes their dynamical behavior. The port variables $\boldsymbol{u}$ and $\boldsymbol{y}$ are conjugate variables in the sense that their duality product defines the power flows exchanged with the environment of the system, for instance, currents and voltages in electrical circuits or forces and velocities in mechanical systems, see [8] for more information on PHS.

Remark 1 *Port-Hamiltonian systems are a generalization of classical Hamiltonian systems but with the capability of including dissipation, input/output ports, and non-local coordinates. Thus, any Hamiltonian systems can be represented by a PHS* (1) *with* $J(\boldsymbol{x}) = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}$, $R = 0$, $G = 0$.

B. Gaussian Process Regression

Let $(\Omega, \mathcal{F}, P)$ be a probability triple with the probability space Ω, the corresponding σ-algebra $\mathcal{F}$ and the probability measure P. Given a mean function $m_{\mathrm{GP}} : \mathbb{R}^n \to \mathbb{R}$ and a kernel function $k : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, a GP $\mathcal{GP}(m_{\mathrm{GP}}, k)$ is a stochastic process on a set $\mathcal{X} \subseteq \mathbb{R}^n$ with marginals defined as follows. Given a sample $f \sim \mathcal{GP}(m_{\mathrm{GP}}, k)$ and a finite collection of points $\boldsymbol{x} \subset \mathcal{X}$, the vector of pointwise evaluations $f(\boldsymbol{x})$ is distributed according to the normal distribution $f(\boldsymbol{x}) \sim \mathcal{N}(m_{\mathrm{GP}}(\boldsymbol{x}), K(\boldsymbol{x},\boldsymbol{x}))$, where $K(\boldsymbol{x},\boldsymbol{x})$ is a matrix of pairwise kernel evaluations at the points in $\boldsymbol{x}$. One of the main advantages of GPs is their combined use with Bayes' Theorem to provide tractable statistical inference for regression [13]. In this regard, consider the output $y \in \mathbb{R}$ of a continuous function $f\colon \mathbb{R}^n \to \mathbb{R}$. The measurements might be affected by Gaussian noise such that $y = f(\boldsymbol{x}) + \eta$ with $\eta \sim \mathcal{N}(0, \sigma^2)$. We place a mean-zero GP prior with kernel k on f. The training set $\mathcal{D} = \{X, Y\}$ denotes the set of input data, $X = \left[\boldsymbol{x}^{\{1\}}, \boldsymbol{x}^{\{2\}}, \ldots, \boldsymbol{x}^{\{N\}}\right] \in \mathbb{R}^{n\times N}$ and measured output data, $Y = \left[y^{\{1\}}, y^{\{2\}}, \ldots, y^{\{N\}}\right]^\top \in \mathbb{R}^N$. For a test input $\boldsymbol{x}^* \in \mathbb{R}^n$, the estimation of $f(\boldsymbol{x}^*)$ is provided by conditioning on the data which leads to the posterior

$$\mu\left(f \,|\, \boldsymbol{x}^*, \mathcal{D}\right) = \boldsymbol{k}\left(\boldsymbol{x}^*, X\right)^\top K^{-1} Y, \tag{2}$$

$$\mathrm{var}\left(f \mid \boldsymbol{x}^*, \mathcal{D}\right) = k\left(\boldsymbol{x}^*, \boldsymbol{x}^*\right) - \boldsymbol{k}\left(\boldsymbol{x}^*, X\right)^\top K^{-1} \boldsymbol{k}\left(\boldsymbol{x}^*, X\right).$$

The kernel k is a measure for the correlation of two inputs $(\boldsymbol{x}, \boldsymbol{x}')$. The function $K\colon \mathbb{R}^{n\times N} \times \mathbb{R}^{n\times N} \to \mathbb{R}^{N\times N}$ is called the Gram matrix whose elements are given by $K_{j',j} = k(X_{:,j'}, X_{:,j}) + \delta(j,j')\sigma^2$ for all $j', j \in \{1, \ldots, N\}$ with the delta function $\delta(j,j') = 1$ for $j = j'$ and zero, otherwise. The vector-valued function $\boldsymbol{k}\colon \mathbb{R}^n \times \mathbb{R}^{n\times N} \to \mathbb{R}^N$, with the elements $k_j = k(\boldsymbol{x}^*, X_{:,j})$ for all $j \in \{1, \ldots, N\}$, expresses the covariance between $\boldsymbol{x}^*$ and the input training data X. The selection of the kernel and the determination of the corresponding hyperparameters can be seen as degrees of freedom of the regression. A popular kernel for GP models of dynamical systems is the squared exponential kernel, which we will use throughout the following sections.

C. Problem Setting

We consider the problem of learning a physical system whose dynamics can be written in Port-Hamiltonian form (1).

[1] Vectors $\boldsymbol{a}$ and vector-valued functions $\boldsymbol{f}(\cdot)$ are denoted with bold characters. Matrices are described with capital letters. I_n is the n-dimensional identity matrix and 0_n the zero matrix. The expression $A_{:,i}$ denotes the i-th column of A. For a positive semidefinite matrix Λ, $\|x-y\|_\Lambda^2 = (x-y)^\top\Lambda(x-y)$. $\mathbb{R}_{>0}$ denotes the set of positive real number whereas $\mathbb{R}_{\geq 0}$ is the set of non-negative real numbers. $\mathcal{C}^i$ denotes the class of i-th times differentiable functions. The operator $\nabla_{\boldsymbol{x}}$ with $\boldsymbol{x} \in \mathbb{R}^n$ denotes $[\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n}]^\top$.

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 14:44:34 UTC from IEEE Xplore. Restrictions apply.

We assume that we have access to (potentially noisy) observations $\tilde{\boldsymbol{x}}(t) \in \mathbb{R}^n$ of the system state $\boldsymbol{x}(t) \in \mathbb{R}^n$ whose evolution over time $t \in \mathbb{R}_{\geq 0}$ follows

$$\dot{\boldsymbol{x}} = [J(\boldsymbol{x}) - R(\boldsymbol{x})]\nabla_{\boldsymbol{x}} H(\boldsymbol{x}) + G(\boldsymbol{x})\boldsymbol{u} \tag{3}$$

starting at $\boldsymbol{x}(0) \in \mathbb{R}^n$. The Hamiltonian $H \in \mathcal{C}^\infty$ is assumed to be *completely unknown* due to unstructured uncertainties in the system typically imposed by nonlinear springs, physical coupling effects, or highly nonlinear electrical and magnetic fields. The parametric structures of the interconnection matrix $J\colon \mathbb{R}^n \to \mathbb{R}^{n\times n}$, dissipation matrix $R\colon \mathbb{R}^n \to \mathbb{R}^{n\times n}$ and I/O matrix $G\colon \mathbb{R}^n \to \mathbb{R}^{n\times m}$ are assumed to be known but the parameters themselves are assumed to be unknown. In detail, the unknown set of parameters will be described by $\boldsymbol{\varphi}_J \in \Phi_J \subseteq \mathbb{R}^{n_{\varphi_J}}, n_{\varphi_J} \in \mathbb{N}$ for the estimated interconnection matrix $\hat{J}(x|\boldsymbol{\varphi}_J) \in \mathbb{R}^{n\times n}$, $\boldsymbol{\varphi}_R \in \Phi_R \subseteq \mathbb{R}^{n_{\varphi_R}}, n_{\varphi_R} \in \mathbb{N}$ for the estimated dissipation matrix $\hat{R}(x|\boldsymbol{\varphi}_R) \in \mathbb{R}^{n\times n}$ and $\boldsymbol{\varphi}_G \in \Phi_G \subseteq \mathbb{R}^{n_{\varphi_G}}, n_{\varphi_G} \in \mathbb{N}$ for the estimated I/O matrix $\hat{G}(x|\boldsymbol{\varphi}_G) \in \mathbb{R}^{n\times m}$.

Problem Formulation: Given a dataset of time-stamps $\{t_i\}_{i=1}^N$ and noisy state observations with inputs, i.e. $\{\tilde{\boldsymbol{x}}(t_i), \boldsymbol{u}(t_i)\}_{i=1}^N$, learn a Hamiltonian $\hat{H}$ and parameters $\boldsymbol{\varphi_J}, \boldsymbol{\varphi_R}, \boldsymbol{\varphi_G}$ for the estimated interconnection, dissipation, and I/O matrix, $\hat{J}, \hat{R}, \hat{G}$, respectively, such that the observed data is described by the PHS

$$\dot{\boldsymbol{x}} = [\hat{J}(\boldsymbol{x}) - \hat{R}(\boldsymbol{x})]\nabla_{\boldsymbol{x}} \hat{H}(\boldsymbol{x}) + \hat{G}(\boldsymbol{x})\boldsymbol{u}. \tag{4}$$

We assume that the data $\tilde{\boldsymbol{x}}(t_i)$ is generated according to $\tilde{\boldsymbol{x}}(t_i) = \boldsymbol{x}(t_i) + \boldsymbol{\eta}$ where $\boldsymbol{x}(t)$ comes from some true PHS (3) with unknown H, unknown parameters of the matrices J, R, G, and the noise $\boldsymbol{\eta} \in \mathbb{R}^n$ is distributed according to a zero-mean Gaussian $\eta \sim \mathcal{N}(\mathbf{0}, \mathrm{diag}[\sigma_1^2, \ldots, \sigma_n^2])$ where the variances $\sigma_1^2, \ldots, \sigma_n^2 \in \mathbb{R}_{\geq 0}$ are unknown.

III. Gaussian Process Port-Hamiltonian Systems

In this section, we propose Gaussian process Port-Hamiltonian systems (GP-PHS) whose structure is visualized in Fig. 2. Starting with data of a physical system (on the left), we model the unknown Hamiltonian as nonparametric, probabilistic function. Additionally, unknown parameters of the interconnection matrix J, the dissipation matrix R and the I/O matrix G can be estimated exploiting the Bayesian nature of the GP. In contrast to parametric models our approach is beneficial in several ways. First, we do not rely on prior knowledge about the parametric structure of the Hamiltonian. For instance, instead of assuming a linear spring model and fitting Hooke's spring constant based on data, the nonparametric structure of the GP allows to learn complex, nonlinear spring models.

Second, the Bayesian nature of the GP enables the model to represent *all possible* PHS under the Bayesian prior based on a finite number of data points. This is not only interesting from a model identification perspective, but this uncertainty quantification can also be explicitly useful for model-based control approaches. However, using a GP to learn the Hamiltonian raises three challenges: (i) The most straightforward application of GP regression, as presented in Section II-B, would require us to have pointwise evaluations of $H(\boldsymbol{x})$ but we do not have access to measurements of the Hamiltonian directly. (ii) The Hamiltonian only appears in the dynamics $\dot{\boldsymbol{x}}$ after being differentiated and we have no direct measurements of $\dot{\boldsymbol{x}}$. (iii) We need to find a kernel such that samples of the GP with this kernel always (or at least almost surely) result in physically correct PHS dynamics.

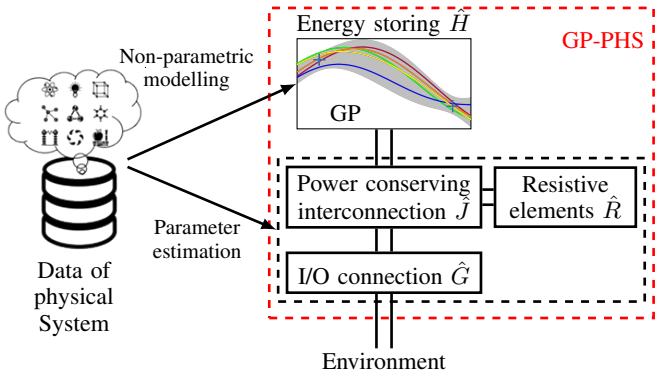


Fig. 2. Block diagram of a Gaussian Process Port-Hamiltonian system.

We addresses these challenges in the following section where we present the general structure of GP-PHS followed by the training and prediction process.

A. Modeling

First, we place a GP prior on the estimated Hamiltonian $\hat{H} \sim \mathcal{GP}(0, k(\boldsymbol{x}, \boldsymbol{x}'))$, where k is the squared exponential kernel. This choice of the kernel results in sampled Hamiltonians which are smooth and possesses a universality property which allows us to approximate any continuous function with it, see [13]. To overcome the fact that state measurements depend only on the partial derivatives of the Hamiltonian and not on the Hamiltonian itself, we leverage the property that GPs are closed under affine operations. Thus, the partial derivative can be included in the GP. The system dynamics are given as an affine transformation of the Hamiltonian in the PH model (4), so we can introduce a modified GP prior

$$\dot{\boldsymbol{x}} \sim \mathcal{GP}(\hat{G}(\boldsymbol{x})\boldsymbol{u}, k_{phs}(\boldsymbol{x}, \boldsymbol{x}'))$$

on the state evolution of (4). Here, we introduce the new PHS kernel $k_{phs}\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^{n\times n}$ given by

$$\begin{aligned} k_{phs}(\boldsymbol{x}, \boldsymbol{x}') &= \sigma_f^2 \hat{J}_R(\boldsymbol{x} \mid \boldsymbol{\varphi}_J, \boldsymbol{\varphi}_R)\Pi(\boldsymbol{x}, \boldsymbol{x}')\hat{J}_R^\top(\boldsymbol{x}' \mid \boldsymbol{\varphi}_J, \boldsymbol{\varphi}_R) \\ \Pi_{i,j}(\boldsymbol{x}, \boldsymbol{x}') &= \frac{\partial}{\partial z_i \partial z_j} \exp(-\|\boldsymbol{z} - \boldsymbol{z}'\|_\Lambda^2)\Big|_{\boldsymbol{z}=\boldsymbol{x}, \boldsymbol{z}'=\boldsymbol{x}'} \end{aligned} \tag{5}$$

where $\Pi\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^{n\times n}$ describes the Hessian of the kernel and $\hat{J}_R(\boldsymbol{x} \mid \boldsymbol{\varphi}_J, \boldsymbol{\varphi}_R) = \hat{J}(\boldsymbol{x} \mid \boldsymbol{\varphi}_J) - \hat{R}(\boldsymbol{x} \mid \boldsymbol{\varphi}_R)$ is used for notational simplification. Note, that the PHS kernel (5) is a matrix kernel as it maps to $\mathbb{R}^{n\times n}$, see [18]. The hyperparameters of k_{phs} are the signal noise $\sigma_f \in \mathbb{R}_{>0}$, the lengthscales $\Lambda = \mathrm{diag}(l_1^2, \ldots, l_n^2) \in \mathbb{R}_{>0}^n$, and the parameter vectors $\boldsymbol{\varphi}_J, \boldsymbol{\varphi}_R, \boldsymbol{\varphi}_G$. Next, we prove that a GP with PHS kernel generates valid samples of a PHS with probability 1.

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 14:44:34 UTC from IEEE Xplore. Restrictions apply.

Proposition 1 *Consider a vector field $\boldsymbol{f}\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ sampled from a GP equipped with the PHS kernel* (5)*, i.e., $\boldsymbol{f} \sim \mathcal{GP}(\hat{G}(\boldsymbol{x})\boldsymbol{u}, k_{phs}(\boldsymbol{x}, \boldsymbol{x}'))$. For all realizations of $\boldsymbol{f}$ in the sample space Ω, the dynamics*

$$\begin{aligned} \dot{\boldsymbol{x}} &= \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}, \omega), \quad \omega \in \Omega \\ \boldsymbol{y} &= \hat{G}(\boldsymbol{x})^\top \nabla_{\boldsymbol{x}} \hat{H}(\boldsymbol{x}, \omega), \end{aligned} \tag{6}$$

describe a Port-Hamiltonian system that is almost surely passive with respect to the supply rate $\boldsymbol{u}^\top \boldsymbol{y}$.

Proof: As we place a GP with squared exponential kernel $k(\boldsymbol{x}, \boldsymbol{x}') = \sigma_f \exp(-\|\boldsymbol{x} - \boldsymbol{x}'\|_\Lambda^2)$ on the estimated Hamiltonian $\hat{H}$, all realizations $\hat{H}(\boldsymbol{x}, \omega)$ with $\omega \in \Omega$ are smooth functions in $\boldsymbol{x}$, see [13, Section 4.2]. Using the fact that every smooth function $\hat{H} : \mathbb{R}^n \to \mathbb{R}$ defines a Port-Hamiltonian vector field under the affine transformation $\hat{H} \mapsto \hat{J}_R \nabla \hat{H} + \hat{G}u$, see [19], all realizations $\hat{H}(\boldsymbol{x}, \omega)$ define Port-Hamiltonian vector fields.

To show passivity, we must first show that there exist a $c_0 \in \mathbb{R}$ such that for almost all $\omega \in \Omega$, $\hat{H}(\boldsymbol{x}, \omega) \geq c_0$, see [7]. This guarantees that infinite energy cannot be extracted from the system. As the mean and covariance of a GP with squared exponential kernel are bounded, see [20], the standard deviation metric $d(\boldsymbol{x}, \boldsymbol{x}') = \mathrm{var}(\hat{H}(\boldsymbol{x}) - \hat{H}(\boldsymbol{x}'))^{1/2}$ for all $\boldsymbol{x}, \boldsymbol{x}' \in \mathbb{R}^n$ is totally bounded, and we get $d(\boldsymbol{x}, \boldsymbol{x}') < c_1$ for some $c_1 \in \mathbb{R}_{>0}$. This, together with the sample path continuity resulting from the squared exponential kernel allows us to appeal to the central limit theorem for stochastic processes, and conclude that $\hat{H}$ is almost surely bounded, i.e, $\mathrm{P}(\sup_{\boldsymbol{x} \in \mathbb{R}^n} |\hat{H}| < \infty) = 1$. Thus, we have shown that there exists a $c_0 \in \mathbb{R}$ such that almost surely $\hat{H} > c_0$. Further, we can now consider the time-derivative of the sampled Hamiltonian for a $\omega \in \Omega$

$$\begin{aligned} \dot{\hat{H}}(\boldsymbol{x}, \omega) &= \nabla_{\boldsymbol{x}}^\top \hat{H}(\boldsymbol{x}, \omega) \hat{J}_R(\boldsymbol{x}) \nabla_{\boldsymbol{x}} \hat{H}(\boldsymbol{x}, \omega) \\ &\quad + \nabla_{\boldsymbol{x}}^\top \hat{H}(\boldsymbol{x}, \omega) \hat{G}(\boldsymbol{x}) \boldsymbol{u} \\ &= -\nabla_{\boldsymbol{x}}^\top \hat{H}(\boldsymbol{x}, \omega) \hat{R}(\boldsymbol{x}) \nabla_{\boldsymbol{x}} \hat{H}(\boldsymbol{x}, \omega) + \boldsymbol{u}^\top \boldsymbol{y} \leq \boldsymbol{u}^\top \boldsymbol{y}, \end{aligned} \tag{7}$$

where $\boldsymbol{u}^\top \boldsymbol{y}$ is the supply rate. As the dissipation matrix $\hat{R}$ is positive semi-definite by definition, equation (7) can be simplified to $\dot{H}(\boldsymbol{x}) \leq \boldsymbol{u}^\top \boldsymbol{y}$. Thus, the change in the system's total energy $\hat{H}$ is less the supply rate with the difference of the dissipation energy. ∎

As consequence of Proposition 1, the PHS kernel allows us to build physically correct models in terms of conversation or dissipation of energy. To model more complex systems, we often wish to combine two or more PHS through an interconnection. It is known that the class of PHS are closed under such interconnections, see [21]. We will show that GP-PHS share the same characteristic.

Proposition 2 *Consider two GP-PHS* (6) *described by $\Sigma(\hat{J}_1, \hat{R}_1, \hat{H}_1, \hat{G}_1)$ with input dimension $m_1 \in \mathbb{N}$ and $\Sigma(\hat{J}_2, \hat{R}_2, \hat{H}_2, \hat{G}_2)$ with input dimension $m_2 \in \mathbb{N}$, respectively. Let $(\boldsymbol{u}_1^c, \boldsymbol{y}_1^c) \in \mathbb{R}^{m_c \times m_c}$ and $(\boldsymbol{u}_2^c, \boldsymbol{y}_2^c) \in \mathbb{R}^{m_c \times m_c}$ be corresponding input/output pairs of dimension m_c with $m_c \leq \min\{m_1, m_2\}$ for the connection of the two GP-PHS. Then, the energy preserving interconnection $\boldsymbol{u}_1^c = -\boldsymbol{y}_2^c$ and $\boldsymbol{u}_2^c = \boldsymbol{y}_1^c$ yields a GP-PHS.*

Proof: First, we start with the definition of two GP-PHS. The first system is given by

$$\begin{aligned} \dot{\boldsymbol{x}} &= \hat{J}_{R1}(\boldsymbol{x}) \nabla_{\boldsymbol{x}} \hat{H}_1(\boldsymbol{x}, \omega_1) + \hat{G}_1(\boldsymbol{x}) \boldsymbol{u}_1 \\ \boldsymbol{y}_1 &= \hat{G}_1(\boldsymbol{x})^\top \nabla_{\boldsymbol{x}} \hat{H}_1(\boldsymbol{x}, \omega_1), \end{aligned}$$

with Hamiltonian $\hat{H}_1 \sim \mathcal{GP}(0, k(\boldsymbol{x}, \boldsymbol{x}'))$, state $\boldsymbol{x} \in \mathbb{R}^{n_1}$, sample $\omega_1 \in \Omega$, $\hat{J}_{R1}(\boldsymbol{x}) = \hat{J}_1(\boldsymbol{x}) - \hat{R}_1(\boldsymbol{x})$, and input/output $\boldsymbol{u}_1, \boldsymbol{y}_1 \in \mathbb{R}^{m_1}$. We separate the I/O matrix $\hat{G}_1$ into $\hat{G}_1^{\mathrm{c}}(\boldsymbol{x}) \in \mathbb{R}^{n_1 \times m_c}$ and $\hat{G}_1^{\mathrm{ex}}(\boldsymbol{x}) \in \mathbb{R}^{n_1 \times m_1 - m_c}$ such that $\hat{G}(\boldsymbol{x})\boldsymbol{u}_1 = \hat{G}_1^{\mathrm{c}} \boldsymbol{u}_1^{\mathrm{c}} + \hat{G}_1^{\mathrm{ex}} \boldsymbol{u}_1^{\mathrm{ex}}$ with external input $\boldsymbol{u}_1^{\mathrm{ex}} \in \mathbb{R}^{m_1 - m_c}$. The output for connection $\boldsymbol{y}_1^{\mathrm{c}}$ is given by $\boldsymbol{y}_1^{\mathrm{c}} = \hat{G}_1^{\mathrm{c}}(\boldsymbol{x})^\top \nabla_{\boldsymbol{x}} \hat{H}_1(\boldsymbol{x}, \omega_1)$.
Analogously, the second system is defined by

$$\begin{aligned} \dot{\boldsymbol{\xi}} &= \hat{J}_{R2}(\boldsymbol{\xi}) \nabla_{\boldsymbol{\xi}} \hat{H}_2(\boldsymbol{\xi}, \omega_2) + \hat{G}_2(\boldsymbol{\xi}) \boldsymbol{u}_2 \\ \boldsymbol{y}_2 &= \hat{G}_2(\boldsymbol{\xi})^\top \nabla_{\boldsymbol{\xi}} \hat{H}_2(\boldsymbol{\xi}, \omega_2), \end{aligned}$$

with Hamiltonian $\hat{H}_2 \sim \mathcal{GP}(0, k(\boldsymbol{\xi}, \boldsymbol{\xi}'))$, state $\boldsymbol{\xi} \in \mathbb{R}^{n_2}$, sample $\omega_2 \in \Omega$, $\hat{J}_{R2}(\boldsymbol{\xi}) = \hat{J}_2(\boldsymbol{\xi}) - \hat{R}_2(\boldsymbol{\xi})$, and input/output $\boldsymbol{u}_2, \boldsymbol{y}_2 \in \mathbb{R}^{m_2}$. The I/O matrix $\hat{G}_2$ is separated into $\hat{G}_2^{\mathrm{c}}(\boldsymbol{\xi}) \in \mathbb{R}^{n_2 \times m_c}$ and $\hat{G}_2^{\mathrm{ex}}(x) \in \mathbb{R}^{n_2 \times m_2 - m_c}$ such that $\hat{G}(\boldsymbol{\xi})\boldsymbol{u}_2 = \hat{G}_2^{\mathrm{c}} \boldsymbol{u}_2^{\mathrm{c}} + \hat{G}_2^{\mathrm{ex}} \boldsymbol{u}_2^{\mathrm{ex}}$ with $\boldsymbol{u}_2^{\mathrm{ex}} \in \mathbb{R}^{m_2 - m_c}$. The output for connection $\boldsymbol{y}_2^{\mathrm{c}}$ is given by $\boldsymbol{y}_2^{\mathrm{c}} = \hat{G}_2^{\mathrm{c}}(\boldsymbol{\xi})^\top \nabla_{\boldsymbol{\xi}} \hat{H}_2(\boldsymbol{\xi}, \omega_2)$.
For the interconnection $\boldsymbol{u}_1^{\mathrm{c}} = -\boldsymbol{y}_2^{\mathrm{c}}$ and $\boldsymbol{u}_2^{\mathrm{c}} = \boldsymbol{y}_1^{\mathrm{c}}$, we get

$$\begin{aligned} \begin{bmatrix} \dot{\boldsymbol{x}} \\ \dot{\boldsymbol{\xi}} \end{bmatrix} &= \underbrace{\begin{bmatrix} \hat{J}_{R1}(\boldsymbol{x}) & -\hat{G}_1^{\mathrm{c}}(\boldsymbol{x})[G_2^{\mathrm{c}}(\boldsymbol{\xi})]^\top \\ \hat{G}_2^{\mathrm{c}}(\boldsymbol{\xi})[\hat{G}_1^{\mathrm{c}}(\boldsymbol{x})]^\top & \hat{J}_{R2}(\boldsymbol{\xi}) \end{bmatrix}}_{J(\boldsymbol{x}, \boldsymbol{\xi}) - R(\boldsymbol{x}, \boldsymbol{\xi})} \begin{bmatrix} \nabla_{\boldsymbol{x}} \\ \nabla_{\boldsymbol{\xi}} \end{bmatrix} \hat{H}(\boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{\omega}) \\ &\quad + \underbrace{[\hat{G}_1^{\mathrm{ex}}(\boldsymbol{x}), \hat{G}_2^{\mathrm{ex}}(\boldsymbol{\xi})]}_{G(\boldsymbol{x}, \boldsymbol{\xi})} \begin{bmatrix} \boldsymbol{u}_1^{\mathrm{ex}} \\ \boldsymbol{u}_2^{\mathrm{ex}} \end{bmatrix} \\ \boldsymbol{y} &= [\hat{G}_1^{\mathrm{ex}}(\boldsymbol{x}), \hat{G}_2^{\mathrm{ex}}(\boldsymbol{\xi})]^\top \begin{bmatrix} \nabla_{\boldsymbol{x}} \\ \nabla_{\boldsymbol{\xi}} \end{bmatrix} \hat{H}(\boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{\omega}) \end{aligned}$$

with $\boldsymbol{\omega} = [\omega_1, \omega_2]^\top$, $J(\boldsymbol{x}, \boldsymbol{\xi}), R(\boldsymbol{x}, \boldsymbol{\xi}) \in \mathbb{R}^{n \times n}, n = n_1 + n_2$, and $G(\boldsymbol{x}, \boldsymbol{\xi}) \in \mathbb{R}^{n \times m}, m = m_1 + m_2 - 2m_c$, and output $\boldsymbol{y} \in \mathbb{R}^n$. If we define the joint Hamiltonian, $\hat{H}\colon \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \to \mathbb{R}$, as $\hat{H}(\boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{\omega}) = \hat{H}_1(\boldsymbol{x}, \omega_1) + \hat{H}_2(\boldsymbol{\xi}, \omega_2)$, then we see that

$$\hat{H} \sim \mathcal{GP}\left(\boldsymbol{0}, \begin{bmatrix} k_1(\boldsymbol{x}, \boldsymbol{x}') & 0 \\ 0 & k_2(\boldsymbol{\xi}, \boldsymbol{\xi}'), \end{bmatrix}\right),$$

where $k_i(\boldsymbol{z}, \boldsymbol{z}') = \sigma_{f,i} \exp(-\|\boldsymbol{z} - \boldsymbol{z}\|_{\Lambda_i}^2)$ denotes the squared exponential kernel for $i = \{1, 2\}$ with hyperparameters $\sigma_{f,1}, \sigma_{f,2} \in \mathbb{R}_{>0}$ and $\Lambda_1 \in \mathbb{R}_{>0}^{n_1 \times n_1}, \Lambda_2 \in \mathbb{R}_{>0}^{n_2 \times n_2}$. ∎

Proposition 2 shows that the negative feedback interconnection of two GP-PHS lead to a GP-PHS again. This is in particular interesting for passivity-based control approaches as shown in, e.g., [22]. Next, we describe the learning and sampling procedure.

B. GP-PHS Training

As indicated in the problem formulation, we consider an observed system trajectory

$$\mathcal{D} = \{(t_1, \tilde{\boldsymbol{x}}(t_1), \boldsymbol{u}(t_1)), \ldots, (t_N, \tilde{\boldsymbol{x}}(t_N), \boldsymbol{u}(t_N))\} \tag{8}$$

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 14:44:34 UTC from IEEE Xplore. Restrictions apply.

of the unknown dynamics (3) corresponding to measured inputs $\{\boldsymbol{u}(t_1), \dots, \boldsymbol{u}(t_N)\}$, where N is the total number of training pairs consisting of a time t_i and a noisy state measurement $\tilde{\boldsymbol{x}}(t_i)$. The first challenge we address is how to extract data of the form $\{\boldsymbol{x}, \dot{\boldsymbol{x}}\}$ from (8) so we may apply GP regression with the PHS kernel to learn the dynamics (4). To obtain the derivative $\dot{\boldsymbol{x}}$, we exploit again that GPs are closed under affine operations [23]. With a differentiable kernel k, we learn n separated GPs on the training sets $\mathcal{D}_j = \{(t_i, \tilde{x}_j(t_i))\}_{i=1,\dots,N}$ with $j = 1, \dots, n$. Thus, one GP for each dimension j of the state $\boldsymbol{x} \in \mathbb{R}^n$ is trained. As in Section II-B, we define an input and output matrix by

$$\begin{aligned} T &= [t_1, \dots, t_N] \in \mathbb{R}^{1\times N} \\ \tilde{X} &= [\tilde{\boldsymbol{x}}(t_1), \dots, \tilde{\boldsymbol{x}}(t_N)]^\top \in \mathbb{R}^{N\times n}. \end{aligned} \tag{9}$$

Using a differentiable kernel k, we obtain the distribution for each element of the *state derivative* $\dot{x}_j \in \mathbb{R}$ by

$$\begin{aligned} \mu(\dot{x}_j \mid t, \mathcal{D}) &= \boldsymbol{k}^{(1)}(t, T)^\top K^{-1} \tilde{X}_{:,j} \\ \operatorname{var}(\dot{x}_j \mid t, \mathcal{D}) &= \boldsymbol{k}^{(1,2)}(t, t) - \boldsymbol{k}^{(1)}(t, T)^\top K^{-1} \boldsymbol{k}^{(1)}(t, T), \end{aligned} \tag{10}$$

where $K_{i,i'} = k(t_i, t_{i'}) + \delta(i, i')\sigma_j^2$ for all $i, i' \in \{1, \dots, N\}$ with the delta function $\delta(i, i') = 1$ for $i = i'$ and zero, otherwise. The term $k^{(l)}$ denotes the derivative of the kernel function k with respect to the l-th argument, i.e., $k^{(1)}(t, T)_i = \frac{\partial}{\partial z} k(z, T_i)\big|_{z=t}$, $k^{(2)}(t, T)_i = \frac{\partial}{\partial z} k(t, z)\big|_{z=T_i}$, and $k^{(1,2)}(t, t) = \frac{\partial^2}{\partial z \partial z'} k(z, z')\big|_{z=t, z'=t}$ for $i = 1, \dots, N$. Further, an estimate of the state $\boldsymbol{x}(t_i)$ based on the noise state measurement $\tilde{\boldsymbol{x}}(t_i)$ is obtained by standard GP regression $\mu(x_j \mid t, \mathcal{D}_j) = \boldsymbol{k}(t, T)^\top K^{-1} \tilde{X}_{:,j}$. Thus, we can create a new dataset $\mathcal{E} = \{(\mu(\boldsymbol{x} \mid t_i, \mathcal{D}), \mu(\dot{\boldsymbol{x}} \mid t_i, \mathcal{D})\}_{i=1,\dots,N}$ that is suitable for the learning of an estimated Hamiltonian $\hat{H}$.

Remark 2 *For the sake of simplicity, we focus here on a single trajectory. However, the above procedure can be repeated for multiple trajectories and the dataset $\mathcal{E}$ can be updated accordingly. In fact, there exist efficient procedures for the updating of GPs with new data, e.g., see [24].*

Further, the dataset is exploited for the estimation of the unknown parameters $\boldsymbol{\varphi}_J \in \Phi_J, \boldsymbol{\varphi}_R \in \Phi_R, \boldsymbol{\varphi}_G \in \Phi_G$ for the matrices $\hat{J}(x|\boldsymbol{\varphi}_J), \hat{R}(x|\boldsymbol{\varphi}_R)$, and $\hat{G}(x|\boldsymbol{\varphi}_G)$ of the PH model (4). First, we write the dataset $\mathcal{E}$ as input and output data analogous to (9) by

$$\begin{aligned} X &= [\mu(\boldsymbol{x} \mid t_1, \mathcal{D}), \dots, \mu(\boldsymbol{x} \mid t_N, \mathcal{D})] \in \mathbb{R}^{n\times N} \\ \dot{X} &= [\mu(\dot{\boldsymbol{x}} \mid t_1, \mathcal{D}), \dots, \mu(\dot{\boldsymbol{x}} \mid t_N, \mathcal{D})]^\top \in \mathbb{R}^{N\times n}. \end{aligned} \tag{11}$$

As the unknown parameters are treated as hyperparameters of the GP, the marginal likelihood is given by $\mathrm{p}(\dot{\boldsymbol{x}}|\boldsymbol{\varphi}, X) = \int \mathrm{p}(\dot{\boldsymbol{x}}|\boldsymbol{f}, X)\,\mathrm{p}(\boldsymbol{f}|\boldsymbol{\varphi}, X) d\boldsymbol{f}$, where $\boldsymbol{\varphi} = [\boldsymbol{\varphi}_J^\top, \boldsymbol{\varphi}_R^\top, \boldsymbol{\varphi}_G^\top, \sigma_f, l_1, \dots, l_n]^\top \in \mathbb{R}^{n_{\varphi_J}+n_{\varphi_R}+n_{\varphi_G}+n+1}$ contains the unknown parameters of the PHS and the kernel parameters $\sigma_f, l_1, \dots, l_n \in \mathbb{R}_+$. With the Gaussian prior $\dot{X}|\boldsymbol{\varphi}, X \sim \mathcal{N}(\mathbf{0}, K_{phs})$ we can compute the negative log marginal likelihood (NLML) of the data.

The matrix $K_{phs} \in \mathbb{R}^{nN\times nN}$ describes the covariance

$$\begin{aligned} K_{phs} = &\begin{bmatrix} k_{phs}(X_{:,1}, X_{:,1}) & \dots & k_{phs}(X_{:,1}, X_{:,N}) \\ \vdots & \ddots & \vdots \\ k_{phs}(X_{:,N}, X_{:,1}) & \dots & k_{phs}(X_{:,N}, X_{:,N}) \end{bmatrix} \\ &+ \begin{bmatrix} I_n \operatorname{var}(\dot{\boldsymbol{x}} \mid t_1, \mathcal{D}) & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & I_n \operatorname{var}(\dot{\boldsymbol{x}} \mid t_N, \mathcal{D}) \end{bmatrix} \end{aligned} \tag{12}$$

based on the matrix kernel k_{phs} as shown in (5). We use the posterior variance (10) of the estimated state derivative data $\dot{X}$ as noise in the covariance matrix (12). This allows us to consider the uncertainty of the estimation in the modelling of the PHS. Then, we can compute the NLML

$$-\log \mathrm{p}(\dot{X}|\varphi, X) \sim \dot{X}_0^\top K_{phs}^{-1} \dot{X}_0 + \log|K_{phs}|, \tag{13}$$

with the mean-adjusted output data $\dot{X}_0 = [[\mu(\dot{\boldsymbol{x}} \mid t_1, \mathcal{D}) - \hat{G}\boldsymbol{u}(t_1)]^\top, \dots, [\mu(\dot{\boldsymbol{x}} \mid t_N, \mathcal{D}) - \hat{G}\boldsymbol{u}(t_N)]^\top]^\top$. Finally, the unknown (hyper)parameters $\boldsymbol{\varphi}$ can be computed by minimization of the NLML (13) via, e.g., a gradient-based method as the gradient is analytically tractable.

C. Prediction

Once the GP model is trained, we can draw samples from the posterior distribution using the joint distribution with mean-adjusted output data at a test states $\boldsymbol{x}^* \in \mathbb{R}^n$

$$\begin{bmatrix} \dot{X}_0 \\ \boldsymbol{f}(\boldsymbol{x}^*) \end{bmatrix} = \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K_{phs} & k_{phs}(X, \boldsymbol{x}^*) \\ k_{phs}(X, \boldsymbol{x}^*)^\top & k_{phs}(\boldsymbol{x}^*, \boldsymbol{x}^*) \end{bmatrix}\right),$$

to obtain a vector field $\boldsymbol{f}$ for the GP-PHS model at these states. However, for numerical integration purposes we will need to be able to access this vector field at an arbitrary number of points at arbitrary locations. An appealing idea is to sample the vector field $\dot{\boldsymbol{x}}$ along a grid of points and interpolate between them to create a callable vector field function for an ode-solver. Unfortunately, we have no guarantee that this interpolation will produce a dynamics function that has a PH structure and respects the energy conservation/dissipation of the system.

To overcome this issue, we propose to sample the estimated Hamiltonian $\hat{H}$ itself and create an interpolation of $\hat{H}$ instead of the vector field $\boldsymbol{f}$. In this case, we can use the fact that $\dot{X}_0$ and $\hat{H}$ are related through the linear transformation $\hat{H} \mapsto \hat{J}_R \nabla \hat{H}$ and form the joint prior

$$\begin{bmatrix} \dot{X}_0 \\ \hat{H}(\boldsymbol{x}^*) \end{bmatrix} = \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K_{phs} & \boldsymbol{k}_{\dot{x}H}(X, \boldsymbol{x}^*) \\ \boldsymbol{k}_{\dot{x}H}(X, \boldsymbol{x}^*)^\top & k_{HH}(\boldsymbol{x}^*, \boldsymbol{x}^*) \end{bmatrix}\right), \tag{14}$$

where the vector functions $\boldsymbol{k}_{\dot{x}H}(X, \boldsymbol{x}^*) \in \mathbb{R}^{nN\times 1}$ and $\boldsymbol{k}_{HH}(\boldsymbol{x}^*, \boldsymbol{x}^*) \in \mathbb{R}_+$ are constructed as shown in Section II-B by the kernels $k_{\dot{x}H}(\boldsymbol{x}, \boldsymbol{x}') = \hat{J}_R(\boldsymbol{x}) \nabla_x k_{HH}(\boldsymbol{x}, \boldsymbol{x}')$ and $k_{HH}(\boldsymbol{x}, \boldsymbol{x}') = \sigma_f^2 \exp(-\|\boldsymbol{x} - \boldsymbol{x}'\|_\Lambda^2)$. We may then sample from the posterior of this distribution after conditioning on values of $\dot{X}_0$ to obtain values for $\hat{H}(\boldsymbol{x}^*)$. After interpolating $\hat{H}$ along these points, the error of the interpolation will not affect the PHS properties as shown in the following.

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 14:44:34 UTC from IEEE Xplore. Restrictions apply.

Corollary 1 *Consider a GP-PHS trained on the dataset* (8) *and a sampled Hamiltonian* $\hat{H}(\cdot,\omega)$ *by* (14) *over a finite set* $X^* \subset \mathbb{R}^n$ *with* $\omega \in \Omega$. *Let* $\hat{H}^*\colon \mathbb{R}^n \to \mathbb{R}$ *be a smooth and bounded function approximator of* $\hat{H}(\cdot,\omega)$. *Then,* $\dot{\boldsymbol{x}} = [\hat{J}(\boldsymbol{x}) - \hat{R}(\boldsymbol{x})]\nabla_x \hat{H}^*(\boldsymbol{x}) + \hat{G}(\boldsymbol{x})\boldsymbol{u}$, $\boldsymbol{y} = \hat{G}(\boldsymbol{x})^\top \nabla_{\boldsymbol{x}} \hat{H}^*(\boldsymbol{x})$ *describes a Port-Hamiltonian system that is passive with respect to the supply rate* $\boldsymbol{u}^\top \boldsymbol{y}$.

Proof: The proof results from Proposition 1. ■

As consequence, we can use any function approximator that generates smooth and bounded functions to interpolate between the state-discrete samples of $\hat{H}(\boldsymbol{x}^*,\omega)$ over all $\boldsymbol{x}^* \in X^*$. Possible approaches are spline interpolation [25] or direct approximation of the GP posterior [26], [27], among others. In Algorithm 1, we summarize the steps to achieve samples of a GP-PHS. The complexity of the algorithm is dominated by the cost of training the GP and the sampling of the Hamiltonian $\hat{H}$, that is $\mathcal{O}(N^3 + N^2 n)$ with respect to the number N of training points and the dimension n of the state $\boldsymbol{x}$. The complexity can be reduced by inducing points methods, as in [28], and efficient sampling strategies as in [26]. As data-driven method, the accuracy of GP-PHS typically increases with the number of training points N. Thus, the choice of N is a trade-off between computational complexity and accuracy of the prediciton. However, as shown in Corollary 1, the model output is physically plausible under all circumstances.

Algorithm 1 Learning of GP-PHS

Input: Trajectory $\mathcal{D} \leftarrow \{(t_i, \tilde{\boldsymbol{x}}(t_i))\}_{i=1,\dots,N}$
Input: Control inputs $\{\boldsymbol{u}(t_i)\}_{i=1,\dots,N}$
Obtain $(\boldsymbol{x}, \dot{\boldsymbol{x}})$ **pairs**
Train n independent GPs with $\mathcal{D}_j$ (9)
Create $\mathcal{E} = \{(\mu(\boldsymbol{x} \mid t_i, \mathcal{D}), \mu(\dot{\boldsymbol{x}} \mid t_i, \mathcal{D})\}_{i=1,\dots,N}$
Obtain GP-PHS model:
Train GP-PHS with $\mathcal{E}$ and $\{\boldsymbol{u}(t_i)\}_{i=1,\dots,N}$
Compute posterior variance $\mathrm{var}(\dot{\boldsymbol{x}} \mid T, \mathcal{D})$
Minimize NLML (13) to estimate $\boldsymbol{\varphi}$
Sampling:
Sample a Hamiltonian $\hat{H} \sim \mathcal{GP}$ over finite set (14)
Compute approximation $\hat{H}^*$ of $\hat{H}$ and simulate GP-PHS

IV. Proof of Concept

We consider the dynamics of an iron ball in the magnetic field of a controlled inductor

$$\dot{\boldsymbol{x}} = \begin{bmatrix} 0 & 1 & 0 \\ -1 & -c|x_2| & 0 \\ 0 & 0 & -1/R \end{bmatrix} \nabla_x H(\boldsymbol{x}) + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} u \qquad (15)$$

with the vertical position of the ball x_1, its momentum x_2 and the magnetic flux x_3. The electric resistance is assumed to be $R = 0.1$ and the drag coefficient $c = 1$. The Hamiltonian is given by $H(\boldsymbol{x}) = \frac{1}{2m}x_2^2 + \frac{1}{2}\frac{x_3^2}{L(x_1)}$, with the mass of the ball $m = 0.1$, and the inductance $L(x_1) = 1/(0.1 + x_1^2)$. Thus, the Hamiltonian establishes the coupling between the mechanical and electrical part of the system. For demonstration of learning the system (15) with GP-PHS, we assume that the true Hamiltonian H and the drag coefficient c *are unknown*. The training set $\mathcal{D}$ consists of data recorded from a single trajectory with initial condition $\boldsymbol{x}_0 = [1, 0, 0]^\top$ and rectangular input signal u as shown in Fig. 3. From this trajectory, we collect every 0.05s a time/state data pair which leads to a dataset of 401 training points with additive measurement noise $\mathcal{N}(0, 0.01^2)$. We use a GP with squared exponential kernel to create a dataset where we minimize the NLML. To obtain the GP-PHS model with the unknown drag parameter $c = 1$, we set $\Phi_J = \mathbb{R}_+$ and minimize the NLML, resulting in $\varphi_J = 0.89$ as estimate for c. We sample 2142 equally distributed points for $\hat{H}$ of the joint distribution (14) over the set $[-0.5, 2] \times [-0.2, 0.2] \times [-3, 5]$ and use spline interpolation to achieve $\hat{H}^*$.

Figure 4 shows the sampling and simulation of five trajectories of the GP-PHS model. Note that we use here an

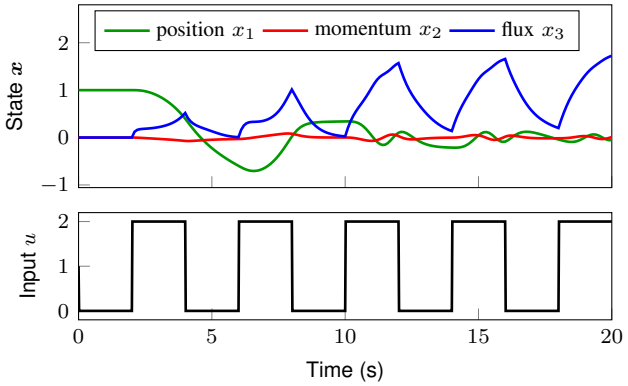


Fig. 3. Single training trajectory over time t with input input u

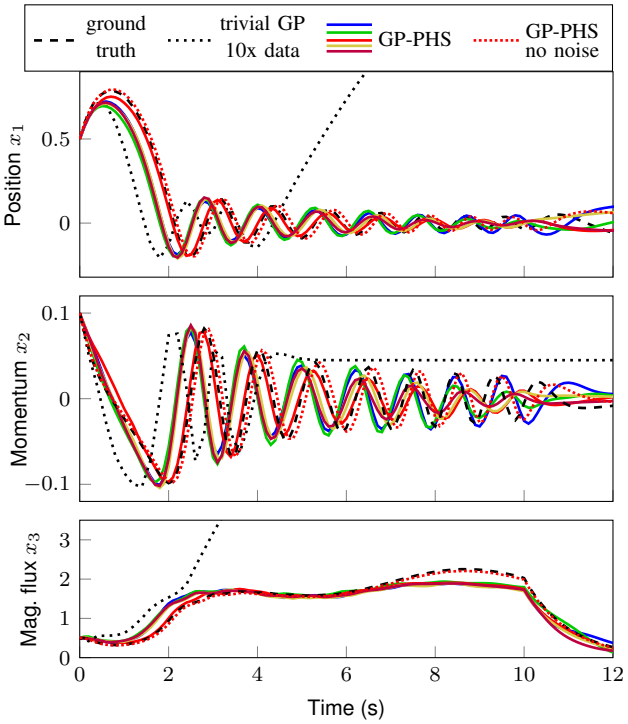


Fig. 4. Five realizations of the GP-PHS and the resulting system trajectories. The realizations for the position x_1, momentum x_2 and the magnetic flux x_3 are significantly better approximations of the true system behavior (dashed line) than a trivial GP learning approach (black dotted).

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 14:44:34 UTC from IEEE Xplore. Restrictions apply.

initial condition $\boldsymbol{x}_0 = [0.5, 0.1, 0.5]^\top$ and input function u which was not seen before in the dataset $\mathcal{D}$. As baseline comparison, we trained a GP function with squared exponential kernel as a direct mapping from $\boldsymbol{x} \mapsto \dot{\boldsymbol{x}}$ without using the underlying PH structure. To achieve a considerable performance, we had to train this naive GP model on a dataset with ten times as many samples as for the GP-PHS, using multiple trajectories with random initial states. However, this baseline approach still does not generalize well as visualized in Fig. 4, where the posterior mean of the trivial GP learning approach (dotted) cannot reproduce the true system behavior. In contrast, the realizations of the proposed GP-PHS generalize well for the new initial condition and input function. For the sake of completeness, we include a sample (red dotted) of a GP-PHS where the measurements are not corrupted by any noise.

Finally, the total energy of the GP-PHS samples are visualized in Fig. 5. All realizations of the GP-PHS represent physical plausible systems as the total energy is non-increasing for a zero input, i.e., for $t > 10$s.

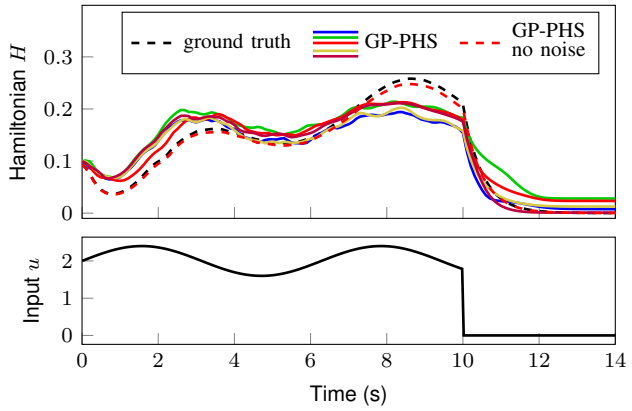


Fig. 5. The Hamiltonians (the total energy in the system) are decreasing over time for a zero input ($t > 10$s) such that all realizations represent a physical correct system behavior.

Conclusion

In this paper, we introduce Gaussian process Port-Hamiltonian systems as a Bayesian learning approach for physical systems. The probabilistic nature of the model allows us to generate all possible realizations of a learned PHS under the GP prior based on a finite dataset. With the proposed PHS kernel, we prove that all realizations of the GP distribution respect the PHS structure. Finally, we show that GP-PHS share the interconnection property and passivity characteristic with PHS. A simulation highlights the superior behavior in contrast to trivial GP regression. In future work, we will develop energy based control strategies based on the proposed GP-PHS model.

References

[1] P. Derler, E. A. Lee, and A. S. Vincentelli, "Modeling cyber–physical systems," *Proceedings of the IEEE*, vol. 100, no. 1, pp. 13–28, 2011.

[2] C. M. Close, D. K. Frederick, and J. C. Newell, *Modeling and analysis of dynamic systems*. John Wiley & Sons, 2001.

[3] E. Andrian, D. Grenier, and M. Rouabhia, "In vitro models of tissue penetration and destruction by porphyromonas gingivalis," *Infection and immunity*, vol. 72, no. 8, pp. 4689–4698, 2004.

[4] Z.-S. Hou and Z. Wang, "From model-based control to data-driven control: Survey, classification and perspective," *Information Sciences*, vol. 235, pp. 3–35, 2013.

[5] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning," *Nature Reviews Physics*, vol. 3, no. 6, pp. 422–440, 2021.

[6] G. Vilasi, *Hamiltonian dynamics*. World Scientific, 2001.

[7] A. Van Der Schaft and D. Jeltsema, "Port-Hamiltonian systems theory: An introductory overview," *Foundations and Trends in Systems and Control*, vol. 1, no. 2-3, pp. 173–378, 2014.

[8] A. Van der Schaft, *L2-gain and passivity techniques in nonlinear control*. Springer, 2000.

[9] S. Greydanus, M. Dzamba, and J. Yosinski, "Hamiltonian neural networks," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[10] T. Bertalan, F. Dietrich, I. Mezić, and I. G. Kevrekidis, "On learning Hamiltonian systems from data," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 29, no. 12, p. 121107, 2019.

[11] S. A. Desai, M. Mattheakis, D. Sondak, P. Protopapas, and S. J. Roberts, "Port-Hamiltonian neural networks for learning explicit time-dependent dynamical systems," *Physical Review E*, vol. 104, no. 3, p. 034312, 2021.

[12] S. P. Nageshrao, G. A. Lopes, D. Jeltsema, and R. Babuška, "Port-Hamiltonian systems in adaptive and learning control: A survey," *IEEE Transactions on Automatic Control*, vol. 61, no. 5, pp. 1223–1238, 2015.

[13] C. E. Rasmussen and C. K. Williams, *Gaussian processes for machine learning*, vol. 1. MIT press Cambridge, 2006.

[14] K. Rath, C. G. Albert, B. Bischl, and U. von Toussaint, "Symplectic Gaussian process regression of maps in Hamiltonian systems," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 31, no. 5, p. 053121, 2021.

[15] S. Ridderbusch, C. Offen, S. Ober-Blöbaum, and P. Goulart, "Learning ode models with qualitative structure using Gaussian processes," in *Proc. of the Conference on Decision and Control*, pp. 2896–2896, 2021.

[16] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Numerical Gaussian processes for time-dependent and nonlinear partial differential equations," *SIAM Journal on Scientific Computing*, vol. 40, no. 1, pp. A172–A198, 2018.

[17] M. A. Bhouri and P. Perdikaris, "Gaussian processes meet NeuralODEs: a Bayesian framework for learning the dynamics of partially observed systems from scarce and noisy data," *Philosophical Transactions of the Royal Society A*, vol. 380, no. 2229, p. 20210201, 2022.

[18] M. A. Alvarez, L. Rosasco, and N. D. Lawrence, "Kernels for vector-valued functions: A review," *Foundations and Trends in Machine Learning*, vol. 4, no. 3, pp. 195–266, 2012.

[19] B. M. Maschke and A. J. van der Schaft, "Port-controlled Hamiltonian systems: modelling origins and systemtheoretic properties," in *Nonlinear Control Systems Design 1992*, pp. 359–365, Elsevier, 1993.

[20] T. Beckers and S. Hirche, "Equilibrium distributions and stability analysis of Gaussian process state space models," in *Proc. of the Conference on Decision and Control*, pp. 6355–6361, 2016.

[21] J. Cervera, A. J. van der Schaft, and A. Baños, "Interconnection of port-Hamiltonian systems and composition of dirac structures," *Automatica*, vol. 43, no. 2, pp. 212–225, 2007.

[22] R. Ortega and E. Garcia-Canseco, "Interconnection and damping assignment passivity-based control: A survey," *European Journal of control*, vol. 10, no. 5, pp. 432–450, 2004.

[23] R. J. Adler, *The geometry of random fields*. SIAM, 2010.

[24] M. F. Huber, "Recursive Gaussian process: On-line regression and learning," *Pattern Recognition Letters*, vol. 45, pp. 85–91, 2014.

[25] C. De Boor and C. De Boor, *A practical guide to splines*, vol. 27. springer-verlag New York, 1978.

[26] J. Wilson, V. Borovitskiy, A. Terenin, P. Mostowsky, and M. Deisenroth, "Efficiently sampling functions from Gaussian process posteriors," in *International Conference on Machine Learning*, pp. 10292–10302, PMLR, 2020.

[27] T. Beckers and S. Hirche, "Prediction with approximated Gaussian process dynamical models," *IEEE Transactions on Automatic Control (to appear)*, 2021.

[28] A. Wilson and H. Nickisch, "Kernel interpolation for scalable structured Gaussian processes (KISS-GP)," in *International conference on machine learning*, pp. 1775–1784, PMLR, 2015.

Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 29,2026 at 14:44:34 UTC from IEEE Xplore. Restrictions apply.