Thermodynamics-Informed Graph Neural Networks

Quercus Hernández, Alberto Badías, Francisco Chinesta, and Elías Cueto

Abstract—In this article, we present a deep learning method to predict the temporal evolution of dissipative dynamic systems. We propose using both geometric and thermodynamic inductive biases to improve accuracy and generalization of the resulting integration scheme. The first is achieved with graph neural networks, which induces a non-Euclidean geometrical prior with permutation-invariant node and edge update functions. The second bias is forced by learning the GENERIC structure of the problem, an extension of the Hamiltonian formalism, to model more general nonconservative dynamics. Several examples are provided in both Eulerian and Lagrangian description in the context of fluid and solid mechanics, respectively, achieving relative mean errors of less than 3% in all the tested examples. Two ablation studies are provided based on recent works in both Physics-informed and geometric deep learning.

Impact Statement—Distilling physical laws from data, either in symbolic form or numeric form, is a problem of utmost importance in the field of Artificial Intelligence. Many different attempts have been developed in recent times based on supervised procedures and different types of neural networks or other more classical regression techniques. In our approach, the scientific knowledge of the physics governing the problem is very important, and is imposed by using the most general formulation: thermodynamics (often referred as the physics of physics). Furthermore, the use of geometric deep learning enables our method to handle unstructured meshed domains, improving its generalizability. With a significant result improvement against previous methods, this technology may contribute to enhance physical simulations in engineering applications such as computational modelling, digital twins or artificial/virtual reality.

Index Terms—Geometric deep learning, graph neural networks (GNNs), learned simulators, Physics-informed deep learning, structure-preserving neural networks, thermodynamics.

I. INTRODUCTION

COMPUTER simulation has become a standard tool in a wide variety of fields, from natural to social sciences, in order to simulate physical phenomena and predict its future behavior. Simulation models enable engineers to find optimal designs which can later be validated in a much more expensive experimental setups. Thus, it is important to develop models based on mathematical and physical foundations which ensure the reliability of the results. With the irruption of the information era, recent research lines have focused on the use of data-based deep learning algorithms which overcome some of the limitations of traditional methods, such as handling highly nonlinear dynamics [1], problems whose mathematical formulation is unknown [2], or real-time simulation performance [3].

However, deep learning is usually very demanding in both data and computational power. A present, challenge is not only to develop numerical algorithms that speed up the already complicated neural network models but also to build smarter architectures which take advantage of the problem structure, potentially reducing the data consumption and improving its generalization. Two historical examples are convolutional [4] and recurrent [5] neural networks, which exploit the structured data in grid elements and timesteps to induce translation and time invariance, respectively, both remaining as major breakthroughs in their respective fields.

Motivated by the study of such architectures, which exploit the invariant quantities of the problem, a new machine learning paradigm recently arose known as geometric deep learning [6]. The key insight is to impose specific constraints related to the symmetries of the problem, acting as a strong inductive bias to the learning process. This framework is not only restricted to regular structured data (images or time-series) but also to a general case of relations (edges) over arbitrary entities (nodes), mathematically represented with a graph.

In a similar way, Physics problems have an intrinsic mathematical structure unveiled by centuries of theoretical and experimental knowledge. From Newton’s laws of motion to the standard model, the laws of nature can often be described as a set of partial differential equations (PDEs), which can also lead neural networks to find the correct solution of a dynamical problem. These are the foundations of the so-called Physics-informed deep learning.

The present work aims to combine both inductive biases in order to learn a physical simulator able to predict the dynamics of complex systems in the context of fluid and solid mechanics.
II. BACKGROUND

A. Physics-Informed Deep Learning

Recent works about predicting Physics with neural networks [1], [7] have demonstrated the convenience of using physical constraints to exploit the dynamical properties of a simulation problem. In fact, the Lagrangian and Hamiltonian reformulations of Newton's classical equations of motion describe a rich mathematical theory based on the Poisson bracket and a symplectic manifold structure. This topological property can be applied to neural networks obtaining robust time integrators, as shown in plenty of recent literature [8]–[12].

This structure can be extended to nonconservative systems using a dissipative bracket with the so-called GENERIC formalism [13], [14]. Several works have already addressed the imposition of this structure bias [3], [15]–[18], but are hard to learn as the standard multilayer perceptron (MLP) architectures may not be adequate for larger systems or phenomena whose GENERIC formulation is more complex.

B. Geometric Deep Learning

The consideration of a general graph theory framework is motivated by the fact that certain data are not structured in the usual Euclidean space, but in a more complex manifold. For instance, solving PDEs on arbitrary domains is a common problem in Physics and engineering, whose solution is rarely found in an Euclidean manifold.

Several graph-based works have achieved great improvements in Physics problems such as predicting atomic bond forces [19], particle tracking in high-energy Physics [20], [21], n-body problem with more general interactions [2], [22], [23], or learning simulators to predict complex fluid interactions [24] and meshed domains [25], [26]. These last approaches include the dynamics of the system by predicting the velocity or acceleration, depending on the order of the governing PDE, and integrating them in time. However, they do not include any physical information about the rest of the predicted variables, which remain as black-box direct predictions.

In this work, we propose a learning method of the GENERIC operators and potentials using graph neural networks (GNNs). This graph-based architecture exploits both the geometrical constraints of discretized systems and the thermodynamically consistent integration scheme of the GENERIC formalism to predict the time evolution of dynamical systems.

The rest of this article is organized as follows. In Section III, we formalize the problem statement and introduce the metriplectic and geometrical inductive biases induced by the proposed architecture. In Section IV, we provide three examples of fluid and solid mechanics comparing the results with other methods. Finally, Section V concludes this article.

III. METHODOLOGY

A. Problem Statement

Let $z \in \mathcal{M} \subseteq \mathbb{R}^n$ be the independent state variables, which fully describe a dynamical system up to a certain level of description, with $\mathcal{M}$ the set of all the admissible states (space state) which is assumed to have the structure of a differentiable manifold in $\mathbb{R}^n$. The given physical phenomenon can be expressed as a system of differential equations encoding the time evolution of its state variables $z$

$$\frac{dz}{dt} = F(z, t), \quad t \in \mathcal{I} = (0, T], \quad z(0) = z_0$$

(1)

where $t$ refers to the time coordinate in the time interval $\mathcal{I}$ and $F(z, t)$ refers to an arbitrary nonlinear differential function.

The goal of this article is to find the convenient mapping $\tilde{F}$ for a dynamical system governed by (1) from data, in order to efficiently predict its time evolution after a prescribed time horizon $T$ using deep learning [27]. The solution is forced to fulfill the basic thermodynamic requirements of energy conservation and entropy inequality restrictions via the GENERIC formalism.

B. Metriplectic Structure: The GENERIC Formalism

In this work, we guarantee the physical meaning of the solution by enforcing the GENERIC structure of the system [13], [14], a generalization of the Hamiltonian paradigm to dissipative systems. The GENERIC formulation of time evolution for nonequilibrium systems, described by the set of $z$ state variables, is given by

$$\frac{dz}{dt} = \{z, E\} + \{z, S\}.$$  

(2)

The conservative contribution is assumed to be of Hamiltonian form, requiring an energy potential $E = E(z)$ and a Poisson bracket $\{z, E\}$ acting on the state vector. This accounts for the reversible dynamics in classical mechanics, expressed by the time structure invariance of the Poisson bracket. Similarly, the remaining irreversible contribution to the energetic balance of the system is generated by the nonequilibrium entropy potential $S = S(z)$ with an irreversible or friction bracket $\{z, S\}$. This term is considered an extension of Landau’s idea of the time evolution of a state variable toward its equilibrium value via a dissipation potential [28].

For practical use, it is convenient to reformulate the bracket notation using two linear operators

$$L: T^*\mathcal{M} \rightarrow \mathcal{T}, \quad M: T^*\mathcal{M} \rightarrow T\mathcal{M}$$

(3)

where $T^*\mathcal{M}$ and $T\mathcal{M}$ represent, respectively, the cotangent and tangent bundles of the state space $\mathcal{M}$. These operators inherit the mathematical properties of the original bracket formulation. The operator $L = L(z)$ represents the Poisson bracket and is required to be skew-symmetric (a symplectic matrix). Similarly, the friction matrix $M = M(z)$ accounts for the irreversible part of the system and is symmetric and positive semi-definite to ensure the positiveness of the dissipation rate.

Replacing the original bracket formulation in (2) with their respective operators, the time-evolution equation for the state variables $z$ is derived

$$\frac{dz}{dt} = L \frac{\partial E}{\partial z} + M \frac{\partial S}{\partial z}.$$  

(4)

However, the restrictions of $L$ and $M$ are not sufficient to guarantee a thermodynamically consistent description of the
dynamics of the system, and two degeneracy conditions are introduced

\[ \{S, z\} = \{E, z\} = 0. \]  

(5)

The first one states that the entropy is a degenerate functional of the Poisson bracket, showing the reversible nature of the Hamiltonian contribution to the dynamics. The second expression states the conservation of the total energy of the system with a degenerate condition of the energy with respect to the friction bracket. These restrictions can be reformulated in a matrix form in terms of the \( L \) and \( M \) operators, resulting in the following degeneracy restrictions:

\[ \frac{\partial S}{\partial z} = M \frac{\partial E}{\partial z} = 0. \]  

(6)

The degeneracy conditions, in addition to the non-negativeness of the irreversible bracket, guarantees the first (energy conservation) and second (entropy inequality) laws of thermodynamics

\[
\begin{align*}
\frac{dE}{dt} &= \{E, E\} = 0 \quad (7a) \\
\frac{dS}{dt} &= \{S, S\} \geq 0. \quad (7b)
\end{align*}
\]

The specific topological requirements based on a pair of skew-symmetric \( (L) \) and symmetric \( (M) \) operators over a smooth manifold is called a metriplectic structure [29], [30]. We make use of this hard constraint to construct a thermodynamically sound integrator, acting as the first inductive bias of our approach. This integration is performed using a forward Euler scheme with time increment \( \Delta t \) and the GENERIC formalism in (4), resulting in the following expression:

\[ z_{t+1} = z_t + \Delta t \left( L \frac{\partial E}{\partial z_t} + M \frac{\partial S}{\partial z_t} \right). \]  

(8)

In order to learn the GENERIC operators \( L, M \), and potentials \( E, S \) for each particle of the domain, we propose the use of graph-based deep learning, which exploits the geometrical structure of that specific domain.

### C. Geometric Structure: Graph Neural Networks

Let \( \mathcal{G} = (V, E, u) \) be a directed graph, where \( V = \{1, \ldots, n\} \) is a set of \( |V| = n \) vertices, \( E \subseteq V \times V \) is a set of \( |E| = e \) edges, and \( u \subseteq \mathbb{R}^{F_v} \) is the global feature vector. Each vertex and edge in the graph is associated with a node and a pairwise interaction between nodes, respectively, in a discretized physical system. The global feature vector defines the properties shared by all the nodes in the graph, such as gravity or elastic properties. For each vertex \( i \in V \), we associate a feature vector \( v_i \in \mathbb{R}^{F_v} \), which represents the physical properties of each individual node. Similarly, for each edge \( (i, j) \in E \), we associate an edge feature vector \( e_{ij} \in \mathbb{R}^{F_e} \).

In practice, the positional state variables of the system \( (q_i) \) are assigned to the edge feature vector \( e_{ij} \) so the edge features represent relative distances \( (q_{ij} = q_i - q_j) \) between nodes, giving a distance-based attentional flavor to the graph network [31]–[33] and translational invariance [34], [35]. The rest of the state variables are assigned to the node feature vector \( v_i \). The external interactions, such as forces applied to the system, are included in an external load vector \( f_i \). A simplified scheme of the graph codification of a physical system is depicted in Fig. 1.

These features are fed into an encode–process–decode scheme [23], which consists on several MLPs shared between all the nodes and edges of the graph. The algorithm consists of five steps (Fig. 2).

1) **Encoding:** We use two MLPs \( (\varphi_v, \varphi_e) \) to transform the vertex and edge initial feature vectors into higher dimensional embeddings \( x_i \in \mathbb{R}^{F_h} \) and \( e_{ij} \in \mathbb{R}^{F_h} \), respectively

\[
\begin{align*}
\varphi_v : \mathbb{R}^{F_v} &\rightarrow \mathbb{R}^{F_h} \\
\varphi_e : \mathbb{R}^{F_e} &\rightarrow \mathbb{R}^{F_h} \\
e_{ij} &\mapsto x_{ij} \\
v_i &\mapsto x_i.
\end{align*}
\]

(9)

2) **Processing:** The processor is the core task of the algorithm, as it shares the nodal information between vertices via message passing and modifies the hidden vectors in order to extract the desired output of the system. First, an MLP \( (\pi_v) \) computes the updated edge features \( x_{ij}' \) for each graph edge, based on the current edge features, global features, and sending and receiving node

\[
\begin{align*}
\pi_v : \mathbb{R}^{3F_h + F_g} &\rightarrow \mathbb{R}^{F_h} \\
(x_{ij}, x_i, x_j, u) &\mapsto x_{ij}'.
\end{align*}
\]

(10)

Then, for each node, the messages are pooled with a permutation invariant function \( \phi \) based on the neighborhood \( \mathcal{N}_i = \{j \in V | (i, j) \in E\} \) of the node \( i \). Last, the node embeddings are updated with a second MLP \( (\pi_v) \) using the current node features, the pooled messages, the external load vector, and the global

\[ v_i' = \varphi_v (\varphi_v (x_i) + \varphi_e (x_{ij}')). \]  

(11)
Fig. 2. Algorithm block scheme used to predict a single-step state variable change in time. (a) The encoder transforms the node and edge features to a learnt embedding. (b) The processor shares the nodal information through the graph via \( M \) message passing modules. (c) The decoder extracts the GENERIC flattened operators and potentials from the processed node embeddings. (d) The reparametrization step builds the symmetries of the \( L \) and \( M \) operators and computes the potential gradients with respect to the network input. (e) The integrator predicts the next time step state variables based on the GENERIC formulation. The whole process is repeated iteratively to get the dynamical rollout of the physical system.

features

\[
\pi_v: \mathbb{R}^{2F_h + F_f + F_y} \longrightarrow \mathbb{R}^{F_h}
\]

\[
(x_i, \phi(x_{ij}'), f_i, u) \mapsto x_i'
\]

where \((\cdot, \cdot)\) denotes the vector concatenation and \(x_i'\) and \(x_{ij}'\) are the updated nodal and edge latent vectors.

The processing step is equivalent to the message passing [36] of 1-step adjacent nodes. In order to get the influence of further graph nodes, the process can be recurrently repeated with both shared or unshared parameters in \( M \) processing blocks and optionally using residual connections [37]. In this approach, we use both unshared parameters and residual connections to each message passing block and sum as aggregation function \( \phi \). Note that the computed messages \( \phi(x_{ij}') \) represent a hidden embedding of the intermolecular interactions of the system (internal messages) whereas the vector \( f_i \) accounts for the external interactions (external messages).

3) Decoding: The last block extracts the relevant physical output information \( y_t \in \mathbb{R}^{F_y} \) of the system from the node latent feature vector, implemented with a MLP (\( \delta_v \)). In this work, we predict for each particle the GENERIC energy \( E \) and entropy \( S \) potentials and the flattened operators \( l \) and \( m \)

\[
\delta_v: \mathbb{R}^{F_h} \longrightarrow \mathbb{R}^{F_y}
\]

\[
x_i' \mapsto y_i = (l, m, E, S).
\]

4) Reparametrization: A last processing step is needed to get the GENERIC parameters before integrating the state variables. Both operators in matrix form \( L \) and \( M \) are constructed using the flattened output of the GNNs \( l \) and \( m \), respectively, reshaped in lower triangular matrices. The skew-symmetric and positive semidefinite conditions are imposed by construction using the following parametrization:

\[
L = l - l^\top, \quad M = mm^\top.
\]

Both \( E \) and \( S \) are directly predicted for every node. Then, these potentials can be differentiated with respect to the network input in order to get the gradients \( \frac{\partial E}{\partial z} \) and \( \frac{\partial S}{\partial z} \) needed for the GENERIC integrator. These gradients are easily obtained using automatic differentiation [38], and ensures the integrability of the energy and entropy gradients [39].

Considering the dimensions of the lower triangular matrices and the scalar value of both potentials, the output dimension of the decoder network is

\[
F_y = \frac{n(n + 1)}{2} + \frac{n(n - 1)}{2} + 1 + 1
\]

where \( n \) represents the dimension of the state variables \( z \).

5) Integration: The single-step integration of the state variables of the system \( z_t \rightarrow z_{t+1} \) is then performed using (8).

D. Learning Procedure

The complete dataset \( D \) is composed by \( N_{\text{sim}} \) multiparametric simulation cases of a dynamical system evolving in time. Each case \( D_i \) contains the labeled pair of a single-step state vector \( z_t \) and its evolution in time \( z_{t+1} \) for each node of the system

\[
D = \{ D_i \}_{i=1}^{N_{\text{sim}}}
\]

\[
D_i = \{ (z_t, z_{t+1}) \}_{t=0}^{T_i}
\]

where the dataset \( D \) is disjointly partitioned in 80\% training, 10\% test, and 10\% validation sets: \( D_{\text{train}}, D_{\text{val}}, \) and \( D_{\text{test}} \) respectively.

The training is performed in a single-snapshot supervision, which has two main advantages: 1) enables parallelization between snapshots, which decreases training time, and 2) avoids intensive memory usage due to a several-snapshot recursive training. The loss function is divided in two terms.

- Data Loss: This term accounts for the correct prediction of the state vector time evolution using the GENERIC integrator. It is defined as the MSE along the graph nodes and state variables between the predicted and the ground-truth time derivative of the state vector in a given snapshot

\[
\mathcal{L}_{\text{data}} = \left\| \frac{dz^\text{GT}}{dt} - \frac{dz^\text{net}}{dt} \right\|_2^2
\]

where \( \| \cdot \|_2 \) denotes the L2-norm. The choice of the time derivative instead of the state vector itself is to regularize the global loss function to a uniform order of magnitude with respect to the degeneracy terms, as shown in (8).

- Degeneracy Loss: This condition is added to the optimization in order to force the degeneracy conditions of the
Poisson and dissipative operators, which ensure thermodynamical consistency of the integrator. It is defined as the MSE along the graph nodes and state variables of two residual terms corresponding to the energy and entropy degeneracy conditions

\[ \mathcal{L}_{\text{deg}}^n = \left\| L \frac{\partial S}{\partial z_n} \right\|^2 + \left\| M \frac{\partial E}{\partial z_n} \right\|^2. \] (17)

Alternative approaches are found in the literature to impose this degeneracy restrictions, such as a specific tensor parametrization of the brackets [17] or forcing orthogonality using additional skew-symmetric matrices [18]. However, we decide to include it as a soft constraint in order to allow more flexibility in the learning process and improve convergence while maintaining the degeneracy conditions up to an admissible error.

The global loss term is a weighted mean of the two terms over the shuffled \( N_{\text{batch}} \) batched snapshots

\[ \mathcal{L} = \frac{1}{N_{\text{batch}}} \sum_{n=0}^{N_{\text{batch}}} \left( \lambda \mathcal{L}_{\text{data}}^n + \mathcal{L}_{\text{deg}}^n \right). \] (18)

As the energy and entropy are supervised only by their gradients, we remark that 1) they are learnt up to an integration constant value and 2) the activation functions must have a sufficient degree of continuity. To meet this second requirement, one must select activations with non-zero second derivative in order to have a correct backpropagation of the weights and biases. Thus, linear or rectified units (ReLU, Leaky ReLU, RReLU) are not appropriate for this task. It is well known [40] that logistic functions such as sigmoid and hyperbolic tangent are universal approximators of any derivative arbitrarily well, but are not optimal for very deep neural networks architectures, as they suffer from several problems such as vanishing gradients. Then, the correct activation functions suitable for learning gradients are the ones which combine both non-zero second derivatives and ReLU-type nonlinearities, such as Softplus, Swish [41], or Mish [42]. In the present work, we use the Swish activation function.

The inputs and outputs of the networks are standardized using the training dataset statistics. Gaussian noise is also added to the inputs during training in order to model the accumulation of error during the time integration [25], which is not contemplated in a single-snapshot training, with the variance of the noise \( \sigma_{\text{noise}}^2 \) as a tunable hyperparameter and zero mean value. All the cases are optimized using Adam [43] and a multistep learning rate scheduler.

The code is fully implemented in Pytorch. Our datasets and trained networks are publicly available online.\(^1\)

### E. Evaluation Metrics

Two ablation studies are performed to evaluate the method presented in this work. The first case is performed using only GNNs with similar architecture and learning procedure used in prior works [24], [25] and no metriplectic integrator. In the second case, we impose the metriplectic structure [15], [16] (from now, SPNN), using standard MLPs with no graph computations. Both alternative methods are tuned for equal parameter count in order to get a fair comparison of the results.

All the results are computed with the integration scheme in (8) iteratively from the initial conditions to the prescribed time horizon \( T \), denoted as rollout. The rollout prediction error is quantified by the relative L2 error, computed with (19) for each snapshot and simulation case

\[ \varepsilon = \frac{\| \mathbf{z}^{GT} - \mathbf{z}^{\text{net}} \|_2}{\| \mathbf{z}^{GT} \|_2}. \] (19)

The results are presented in Figs. 9–11, showing the rollout statistics for all the snapshots divided in train and test simulations, state variables, and method used (ours, GNN, or SPNN).

### IV. Numerical Experiments

#### A. Couette Flow of an Oldroyd-B Fluid

1) **Description:** The first example is a shear (Couette) flow of an Oldroyd-B fluid model (Fig. 3). This is a constitutive model for viscoelastic fluids, considering linear elastic dumbbells as a proxy representation of polymeric chains immersed in a solvent.

\[ \varepsilon = \frac{\| \mathbf{z}^{GT} - \mathbf{z}^{\text{net}} \|_2}{\| \mathbf{z}^{GT} \|_2}. \] (19)

The results are presented in Figs. 9–11, showing the rollout statistics for all the snapshots divided in train and test simulations, state variables, and method used (ours, GNN, or SPNN).

---

\(^1\)[Online]. Available: https://github.com/quercushernandez

---

Fig. 3. Couette flow in an Oldroyd-B fluid. The simulations span different Reynolds and Weissenberg numbers to obtain different flow profiles with a fixed lid velocity.

Fig. 4. Viscoelastic beam problem with a load case. The load position and direction are modified on each simulation, obtaining different stress fields.

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
Fig. 5. (a)–(c) Representation of a snapshot of three test simulations, i.e., not seen by the network on training, of the bending beam problem. (c)–(e) Their respective ground-truth simulations. The color code represents the $xx$ component of the dimensionless stress tensor, scaled $\times 0.001$.

Fig. 6. Unsteady flow past a cylinder obstacle. The flow velocity and cylinder obstacle position are varied to obtain different Reynolds numbers and flow profiles.

The state variables chosen are the position of the fluid on each node of the mesh $q$, its velocity $v$ in the $x$ direction, internal energy $e$, and the conformation tensor shear component $\tau$

$$S = \{ z = (q, v, e, \tau) \in \mathbb{R}^2 \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \}. \quad (20)$$

The edge feature vector contains the relative position of the nodes, whereas the rest of the state variables are part of the node feature vector. An additional one-hot vector $n$ is added to the node features in order to represent the boundary and fluid nodes. The global feature vector $u$ represents the Weissenberg and Reynolds numbers of each simulation, resulting in the following feature vectors:

$$e_{ij} = (q_i - q_j, \|q_{ij}\|_2), \quad v_i = (v, e, \tau, n), \quad u = (Re, We). \quad (21)$$

2) Database and Hyperparameters: The training database for the Couette flow is generated with the CONNFFESSIT technique [44], based on the Fokker–Plank equation [45], using a Monte–Carlo algorithm. The fluid is discretized in the vertical direction with $N_e = 100$ elements and $N = 101$ nodes in a total height of $H = 1$. A total of 10 000 dumbells are considered at each nodal location in the model. The lid velocity is set to $V = 1$, with variable Weissenberg $We \in [1, 2]$ and Reynolds number $Re \in [0, 1]$, summing a total of $N_{\text{sim}} = 100$ cases. The simulation is discretized in $N_T = 150$ time increments of $\Delta t = 6.7 \cdot 10^{-3}$.

Following (21), the dimensions of the graph feature vectors are $F_e = 3$, $F_v = 5$, and $F_\tau = 2$. The hidden dimension of the node and edge latent vectors is $F_h = 10$. The learning rate is set to $l_r = 10^{-3}$ with decreasing order of magnitude on epochs 2000 and 4000, and a total number of $N_{\text{epoch}} = 6000$. The training noise variance is set to $\sigma_{\text{noise}}^2 = 10^{-2}$.

3) Results: The rollout results for the Couette flow are presented in Fig. 9. A substantial improvement is shown in the present approach over the two other methods, which remains in a similar performance. Note that the skewed distributions toward higher errors on each box are due to the error accumulation on snapshots further in time from the starting conditions, where errors are lower. Fig. 8 (left) shows that the
HERNÁNDEZ et al.: THERMODYNAMICS-INFORMED GRAPH NEURAL NETWORKS

Fig. 7. (a)–(c) Representation of a snapshot of three test simulations, i.e., not seen by the network on training, of the cylinder flow problem. (c)–(e) Their respective ground-truth simulations. The color code represents the $x$ component of the dimensionless velocity field.

Fig. 8. Conservation of energy and nondecreasing entropy potentials for a test case of the (a) Couette flow, (b) bending beam, and (c) cylinder flow. Both quantities are averaged across all graph nodes for visualization.

degeneracy conditions imposed by our method ensure the thermodynamical consistency of the learnt energy and entropy potentials.

B. Viscoelastic Bending Beam

1) Description: The next example is a viscoelastic cantilever beam subjected by a bending force (see Fig. 4). The material is characterized by a single-term polynomial strain energy potential, described by the following equation:

$$U = C_{10} (\bar{T}_1 - 3) + C_{01} (\bar{T}_2 - 3) + \frac{1}{D_1} (J_{el} - 1)^2$$  \hspace{1cm} (22)

where $U$ is the strain energy potential, $J_{el}$ is the elastic volume ratio, $\bar{T}_1$ and $\bar{T}_2$ are the two invariants of the left Cauchy–Green deformation tensor, $C_{10}$ and $C_{01}$ are shear material constants, and $D_1$ is the material compressibility parameter. The viscoelastic component is described by a two-term Prony series of the dimensionless shear relaxation modulus

$$g_R(t) = 1 - \bar{g}_1 \left( 1 - e^{\frac{-t}{\tau_1}} \right) - \bar{g}_2 \left( 1 - e^{\frac{-t}{\tau_2}} \right)$$  \hspace{1cm} (23)

with relaxation coefficients of $\bar{g}_1$ and $\bar{g}_2$, and relaxation times of $\tau_1$ and $\tau_2$.

The state variables for the viscoelastic beam on each node are the position $q$, velocity $v$, and stress tensor $\sigma$

$$S = \{ z = (q, v, \sigma) \in \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}^6 \}.$$  \hspace{1cm} (24)

The relative deformed position is included into the edge feature vector, whereas the rest of the variables are part of the node feature vector. An additional one-hot vector $n$ is added to the node features in order to represent the encastre and beam nodes. The external load vector $F$ is included in the node processor MLP as an external interaction. No global feature vector is needed in this case, resulting in the following feature vectors:

$$e_{ij} = (q_i - q_j, \| q_{ij} \|_2), \quad v_i = (v, \sigma, n).$$  \hspace{1cm} (25)

2) Database and Hyperparameters: The prismatic beam dimensions are $H = 10$, $W = 10$, and $L = 40$, discretized in $N_e = 500$ hexahedral linear brick elements and $N = 756$ nodes. The material hyperelastic and viscoelastic parameters are $C_{10} = 1.5 \times 5$, $C_{01} = 5 \times 3$, $D_1 = 10^{-7}$ and $\bar{g}_1 = 0.3$, $\bar{g}_2 = 0.49$, $\tau_1 = 0.2$, and $\tau_2 = 0.5$, respectively. A distributed load
of $F = 10^5$ is applied in $N_{\text{sim}} = 52$ different positions with an orientation perpendicular to the solid surface. The quasi-static simulation is discretized in $N_T = 20$ time increments of $\Delta t = 5 \cdot 10^{-2}$.

Following (25), the dimensions of the graph feature vectors are $F_e = 4$, $F_v = 11$, and $F_g = 0$. The hidden dimension of the node and edge latent vectors is $F_h = 50$. The learning rate is set to $l_r = 10^{-4}$ with decreasing order of magnitude on epochs 600 and 1200, and a total number of $N_{\text{epoch}} = 1800$. The training noise variance is set to $\sigma_{\text{noise}}^2 = 10^{-5}$.

3) Results: The rollout results for the bending viscoelastic beam are presented in Fig. 10. The errors achieved by the present approach are again below the other two methods. The beam-deformed configuration of three different test simulation snapshots is represented in Fig. 5, with the color code representing the $xx$ component of the stress tensor. Similarly to the previous case, Fig. 8 (center) shows the thermodynamical consistency of our dynamical integration.

C. Flow Past a Cylinder

1) Description: The last example consists of a viscous unsteady flow past a cylinder obstacle (see Fig. 6). The flow conditions are set to obtain varying Reynolds regimes, which result in Kármán vortex street and therefore a periodic behavior in the steady state.
The state variables for the flow past a cylinder are the velocity \( \mathbf{v} \) and the pressure field \( P \):

\[
S = \{ z = (\mathbf{v}, P) \in \mathbb{R}^2 \times \mathbb{R} \}. \tag{26}
\]

The flow is computed with an Eulerian description of the output fields. Thus, the nodal coordinates \( (q^0_i) \) are fixed in space and considered as edge features, whereas the whole state variables are assigned to the node features. An additional one-hot vector \( n \) is added to the node features in order to represent the inlet/outlet, walls or fluid nodes. No global feature vector is needed in this case, resulting in the following feature vectors:

\[
e_{ij} = (q^0_i - q^0_j, \|q^0_i\|_2), \quad v_i = (\mathbf{v}, P, n). \tag{27}
\]

2) Database and Hyperparameters: The ground-truth simulations are computed by solving the 2-D Navier Stokes equations. Six different obstacle positions are simulated with varying fluid discretization, which consist of approximately \( N_c = 1100 \) quadrilateral elements and \( N = 1200 \) nodes. No-slip conditions are forced in the stream walls and the cylinder obstacle. The fluid has a density of \( \rho = 1 \) and a dynamic viscosity of \( \mu = 10^{-3} \). The variable freestream velocity is contained within the interval \( v \in [1, 2] \), summing a total of \( N_{\text{sim}} = 30 \) cases. The unsteady simulation is discretized in \( N_T = 300 \) time increments of \( \Delta t = 10^{-2} \).

Following (27), the dimensions of the graph feature vectors are \( F_c = 3, F_v = 8, \) and \( F_n = 0 \). The hidden dimension of the node and edge latent vectors is \( F_h = 128 \). The learning rate is set to \( l_\alpha = 10^{-5} \) with decreasing order of magnitude on epochs 600 and 1200, and a total number of \( N_{\text{epoch}} = 2000 \). The training noise variance is set to \( \sigma^2_{\text{noise}} = 4 \cdot 10^{-4} \).

3) Results: The rollout results for the flow past a cylinder problem are presented in Fig. 11. In this example, the domain varies significantly, using a different unstructured mesh for each simulation. Thus, the graph-based architectures outperform the vanilla SPNN, which is meant for fixed structured problems. Considering the other two methods, our approach outperforms the standard GNN architecture due to the metriplectic structure imposition over the dynamical problem, as depicted in Fig. 8. A single snapshot of the whole rollout of three different test simulations is represented in Fig. 7, with the color code representing the \( x \) component of the velocity field.

V. CONCLUSION

We have presented a method to predict the time evolution of an arbitrary dynamical system based on two inductive biases. The metriplectic bias ensures the correct thermodynamic structure of the integrator based on the GENERIC formalism, whose operators and potentials are estimated using computations over graphs, i.e., exploiting the geometric structure of the problem. The results show relative mean errors of less than 3% in all the tested examples, outperforming two other state-of-the-art techniques based on only Physics-informed and geometric deep learning, respectively. These results confirm that both biases are necessary to achieve higher precision in the predicted simulations. The use of both techniques combine the computational power of geometric deep learning with the rigorous foundation of the GENERIC formalism, which ensures the thermodynamical consistency of the results.

The limitations of the presented technique are related to the computational complexity of the model. Large simulations with fine grids require a high amount of message passing to get the information across the whole domain, or a very fine time discretization, which both result in a high computational cost. Similarly, high-speed phenomena in relation to the wave velocity of the medium might be impossible to model.

Future work may overcome the stated limitations by combining graph representations with model order reduction techniques, such as autoencoders or U-net architectures [46], [47]. The idea is to replace deep message passing with various coarse-graining steps, allowing the boundary information to reach every node in the simulation domain while reducing the number of parameters of the neural network. Another interesting topic to extend our work is to improve the generalization and decrease the amount of training data via equivariant architectures [48], [49], which avoid data augmentation by exploiting the invariance to certain groups such as rotations \( SO(3) \) or general Euclidean transformations \( E(3) \). As the present work is only limited to \textit{in silico} experiments, future work may extend the proposed method to measured datasets in real-world applications, such as digital twins of industrial processes or real-time augmented/virtual reality environments.

REFERENCES

[1] M. Raissi, P. Perdikaris, and G. E. Karniadakis, “Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations,” J. Comput. Phys., vol. 378, pp. 686–707, 2019.

[2] M. Cranmer et al., “Discovering symbolic models from deep learning with inductive biases,” Adv. Neural Inf. Process. Syst., vol. 33, pp. 17429–17442, 2020.

[3] B. Moya, A. Badasia, D. Gonzalez, F. Chinesta, and E. Cueto, “Physics perception in sloshing scenes with guaranteed thermodynamic consistency,” IEEE Trans. Pattern Anal. Mach. Intell., to be published, doi: 10.1109/TPAMI.2021.3160100.

[4] Y. LeCun, P. Haffner, L. Bottou, and Y. Bengio, “Object recognition with gradient-based learning,” in Shape, Contour and Grouping in Computer Vision. Berlin, Germany: Springer, 1999, pp. 319–345.

[5] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning internal representations by error propagation,” Tech. Rep., California Univ. San Diego La Jolla Inst for Cognitive Science, 1985.

[6] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, “Geometric deep learning: Going beyond euclidean data,” IEEE Signal Process. Mag., vol. 34, no. 4, pp. 18–42, Jul. 2017.

[7] M. Raissi and G. E. Karniadakis, “Hidden physics models: Machine learning of nonlinear partial differential equations,” J. Comput. Phys., vol. 357, pp. 125–141, 2018.

[8] S. J. Greydanus, M. Dzamba, and J. Yosinski, “Hamiltonian neural networks,” Adv. Neural Inf. Process. Syst., 2019.

[9] A. Sanchez-Gonzalez, V. Bapst, K. Cranmer, and P. Battaglia, “Hamiltonian graph networks with ode integrators,” 2019, arXiv:1909.12790.

[10] M. Matthaeakis, D. Sondak, A. S. Doga, and P. Protopapas, “Hamiltonian neural networks for solving differential equations,” 2020, arXiv:2001.11107.

[11] S. Desai, M. Matthaeakis, D. Sondak, P. Protopapas, and S. Roberts, “Port-Hamiltonian neural networks for learning explicit time-dependent dynamical systems,” Phys. Rev. E, vol. 104, 2021, Art. no. 033412.

[12] Z. Chen et al., “Symplectic recurrent neural networks,” Int. Conf. Learn. Representations, 2019.

[13] H. C. ottinger and M. Grmela, “Dynamics and thermodynamics of complex fluids. II. Illustrations of a general formalism,” Phys. Rev. E, vol. 56, no. 6, 1997, Art. no. 6633.

[14] M. Grmela and H. C. ottinger, “Dynamics and thermodynamics of complex fluids. I. Development of a general formalism,” Phys. Rev. E, vol. 56, no. 6, 1997, Art. no. 6620.
Quercus Hernández received the B.S. degree in mechanical engineering, in 2014, and the M.Eng. degree in industrial engineering, in 2016, from the University of Zaragoza, Zaragoza, Spain, where he is currently working toward the Ph.D. degree in mechanical engineering. His thesis is oriented toward physics-informed deep learning applied to computational mechanics.

Alberto Badia received the B.Sc. degree in mechanical engineering, in 2011, the M.S. degree in industrial engineering (industrial automation and robotics), in 2016, the M.S. degree in biomedical engineering, in 2020, and the Ph.D. degree from the University of Zaragoza, Zaragoza, Spain, in 2020. He is currently an Assistant Professor with the Polytechnic University of Madrid, Madrid, Spain. He has been working in the area of computer vision and robotics developing 3-D reconstructions and new image processing tools and is working currently in physics-informed learning methods in applied mechanics and bioengineering.

Francisco Chinea received the B.Sc. and M.Sc. degrees at the Polytechnic University of Valencia, Spain, in 1990. He is currently a Professor of Computational Mechanics with the PIMM Laboratory, ENSAM ParisTech, Paris, France, where he was AIRBUS Chair Professor from 2008 to 2012, and has been ESI Chair Professor since 2013. He has authored more than 250 papers in peer-reviewed journals. His research contributions include the proposal and development of efficient model reduction strategies based on the proper generalized decomposition and more particularly the development of separated representations-based strategies. Dr. Chinea is an Associate Member of the University of Wales Institute of Non-Newtonian Fluid Mechanics, a Fellow of the “Institut Universitaire de France” (IUF), and a Fellow of the Spanish Royal Academy of Engineering. He has received many international scientific prizes and awards.

Elias Cueto received the B.Sc. and M.Sc. degrees in the University of Oviedo, Spain, in 1997. He is currently a Professor of Continuum Mechanics with the University of Zaragoza, Zaragoza, Spain. He has worked in model order reduction techniques and real-time simulation for computational surgery and augmented reality applications. His research interests include the development of advanced numerical strategies for complex phenomena. Dr. Cueto has received the J. C. Simo Award of the Spanish Society of Numerical Methods in Engineering, the European Scientific Association of Metal Forming (EASAFORM) Scientific Prize, and the O.C. Zienkiewicz Prize of the European Community on Computational Methods in Applied Sciences, ECCOMAS, among others. He is a Fellow of the EAMBES society.