Deep Energy-Based Modeling of Discrete-Time Physics

Takashi Matsubara*, Ai Ishikawa†, Takaharu Yaguchi†
***Osaka University, †Kobe University**

Abstract: We propose a deep energy-based physical model that admits a specific differential geometric structure. From this structure, the conservation or dissipation law of energy and the mass conservation law follow naturally. To ensure the energetic behavior in discrete time, we also propose an automatic discrete differentiation algorithm that enables neural networks to employ the discrete gradient method.

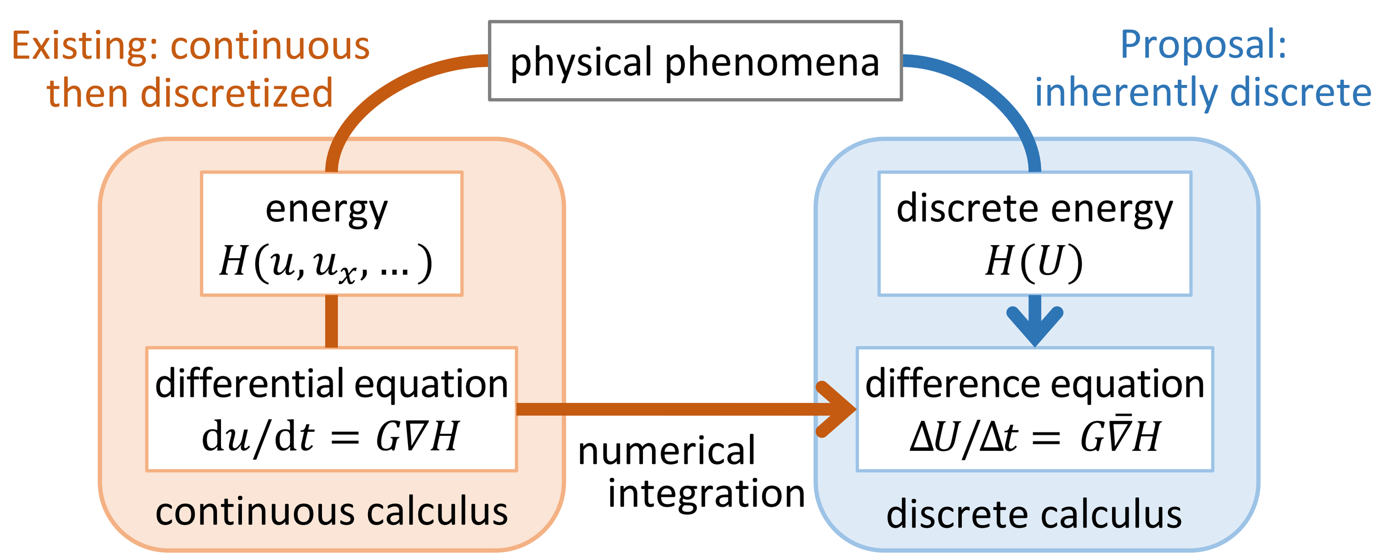
NEURAL INFORMATION PROCESSING SYSTEMS

How Deep Learning Ensures the Laws of Physics?

Goal: a deep neural network that
- approximates physical phenomena from observations.
- ensures the energy conservation and dissipation laws *in discrete time*.

Background:
- The Hamiltonian mechanics is expressed by an ordinary differential equation (ODE) associated with a symplectic structure, which admits the energy conservation law in continuous time [1].
- For computer simulations, the time is discretized and interpolated by a numerical integrator; then, the energy conservation law is destroyed.
- Symplectic integrators preserve an approximated symplectic structure and conserve a modified energy [2], not the original energy.
- Symplectic integrators are not applicable to dissipative systems.
- Systems expressed by partial differential equations (PDEs) have been out of scope of most previous studies.

Proposal:
- generalizes the Hamiltonian mechanics to energy-based modeling.
- modeled in discrete time by the discrete gradient method, which ensures the energetic property in discrete time.



Method: Energy-Based Modeling

The unified form $du/dt = G\nabla H(u)$
- for the state u and the system energy H.
- expresses most physical phenomena.
- exhibits the energetic property determined by the coefficient matrix G.

	ODEs $u = (q, p)$	PDEs $u = (u_1, \dots, u_n)$
Conservative systems G: skew-symmetric $dH/dt = \nabla H^\top G \nabla H = 0$	Hamiltonian mechanics $S = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$ ex. molecular dynamics, astronomical system	Hamiltonian field theory $D = \frac{1}{2\Delta x}\begin{pmatrix} 0 & 1 & & & -1 \\ -1 & 0 & 1 & & \\ & & \ddots & & \\ & & & -1 & 0 & 1 \\ 1 & & & -1 & 0 \end{pmatrix}$ ex. nonlinear wave equation, shallow water wave
Dissipative systems G: negative semidefinite $dH/dt = \nabla H^\top G \nabla H \le 0$	Mechanics with friction $S + R = \begin{pmatrix} 0 & I \\ -I & r \end{pmatrix}, r_k \le 0$ ex. pendulum, real robots	Landau free-energy theory $D_2 = \frac{1}{\Delta x^2}\begin{pmatrix} -2 & 1 & & & 1 \\ 1 & -2 & 1 & & \\ & & \ddots & & \\ & & 1 & -2 & 1 \\ 1 & & & 1 & -2 \end{pmatrix}$ ex. phase separation, cracks

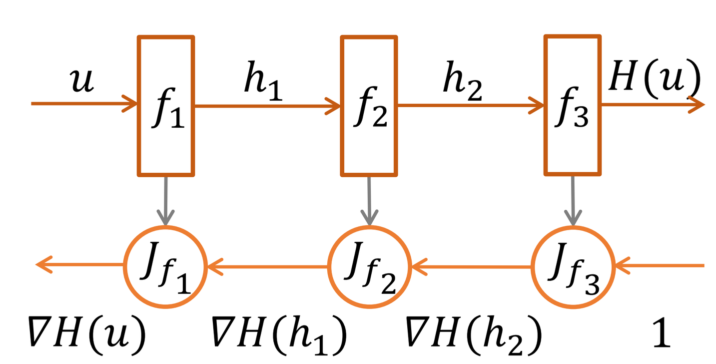
Method: Modeled in Discrete Time

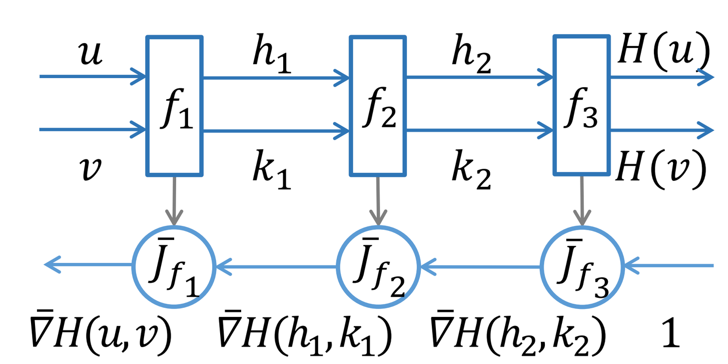
Discrete calculus [3]
- ensures the energy conservation and dissipation laws in discrete time.
- for $\Delta u = u_{n+1} - u_n$ and $\Delta H = H(u_{n+1}) - H(u_n)$.

Continuous calculus		**Discrete calculus**
differential dH		discrete differential $\bar{d}H$
gradient ∇H		discrete gradient $\bar{\nabla} H$
$dH(u) = \nabla H(u) \cdot du$	chain-rule	$\Delta H(u_{n+1}, u_n) = \bar{\nabla} H(u_{n+1}, u_n) \cdot \Delta u$
$dH(ax; u) = a\,dH(x; u)$	linearity	$\bar{d}H(ax; u, v) = a\bar{d}H(x; u, v)$
Continuous-time system		**Discrete-time system**
$du/dt = G\nabla H(u)$		$\Delta u/\Delta t = \bar{G}\bar{\nabla} H(u_{n+1}, u_n) \cdots (*)$
Energetic property		**Discrete energetic property**
$dH/dt = \nabla H \cdot du/dt = \nabla H^\top G \nabla H$		$\Delta H/\Delta t = \bar{\nabla} H \cdot \Delta u/\Delta t = \bar{\nabla} H^\top \bar{G} \bar{\nabla} H$

Automatic discrete differentiation
- The discrete gradient is not unique, and existing ones are inapplicable.
- The new algorithm obtains the discrete gradient of a neural network.

Automatic differentiation		**Automatic discrete differentiation**
$\nabla(f \circ g)(u) = J_g(u)^\top \nabla f(g(u))$		$\bar{\nabla}(f \circ g)(u, v) = \bar{J}_g(u, v)^\top \bar{\nabla} f(g(u), g(v))$
$J_g = W$	linear	$\bar{J}_g = W$
$(J_g)_{kk} = \frac{d\sigma(u^{(k)})}{du^{(k)}}$	nonlinear	$(\bar{J}_g)_{kk} = \frac{\sigma(u^{(k)}) - \sigma(v^{(k)})}{u^{(k)} - v^{(k)}}$





Experiments

At the training phase:
- The objective function is the error in the numerical scheme $(*)$.
- Given two time steps u_n, u_{n+1}, a single back-propagation is enough.

Comparative methods:
- The continuous model is solved by a Runge-Kutta method, the Dormand-Prince method with adaptive time-stepping [4], which divides a time step into many substeps and requires a proportional computational cost.
- To ODE systems with the separable energy function, a symplectic integrator (the leapfrog integrator) is applicable with two neural networks [2].

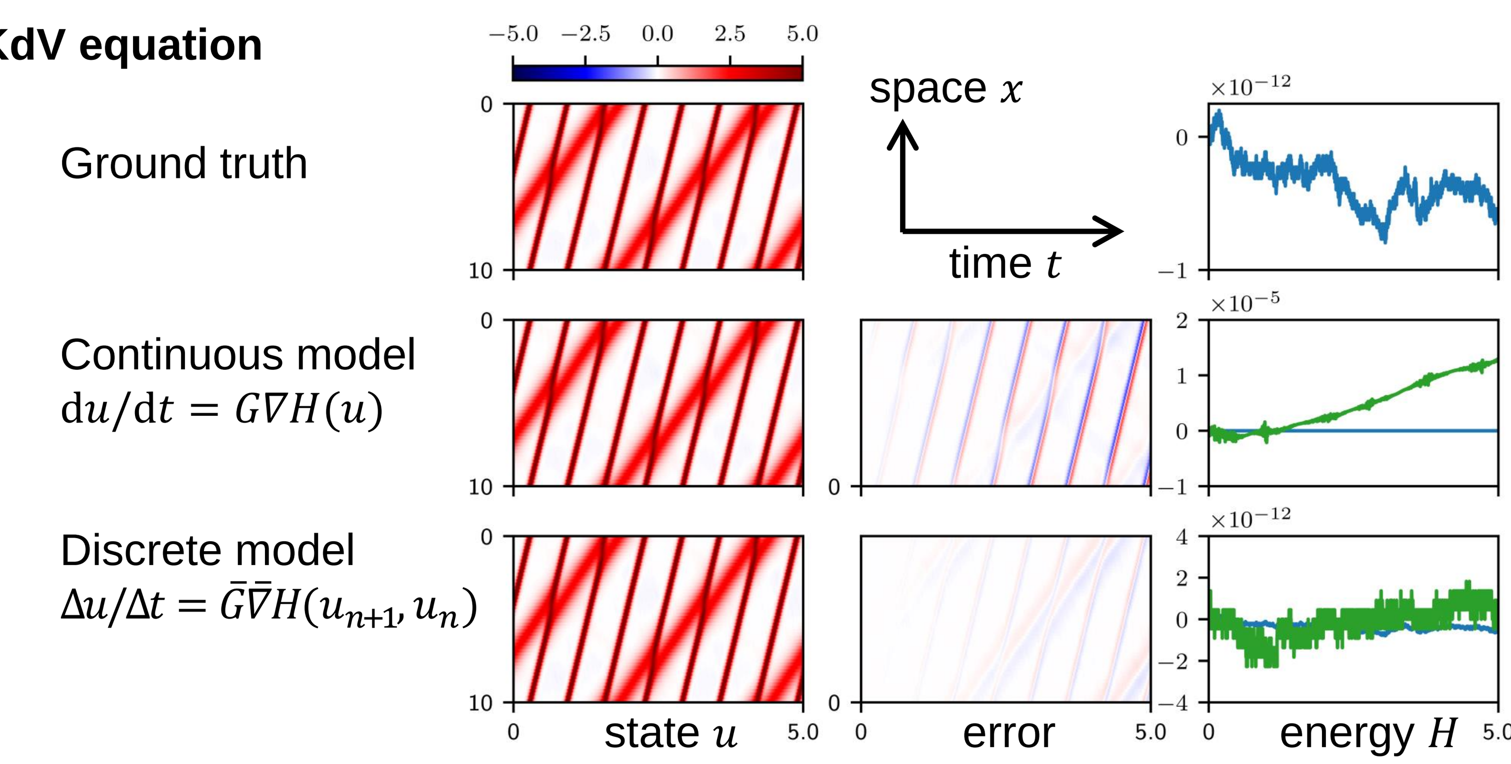
References
[1] Greydanus *et al*. (2019). "Hamiltonian Neural Networks." In: Advances in Neural Information Processing Systems.
[2] Chen *et al*. (2020). "Symplectic Recurrent Neural Networks." In: International Conference on Learning Representations.
[3] Furihata *et al*. (2010). Discrete Variational Derivative Method, Chapman and Hall/CRC.
[4] Chen *et al*. (2018). "Neural Ordinary Differential Equations." In: Advances in Neural Information Processing Systems.

Results

PDE systems (MSEs in appropriate scales)

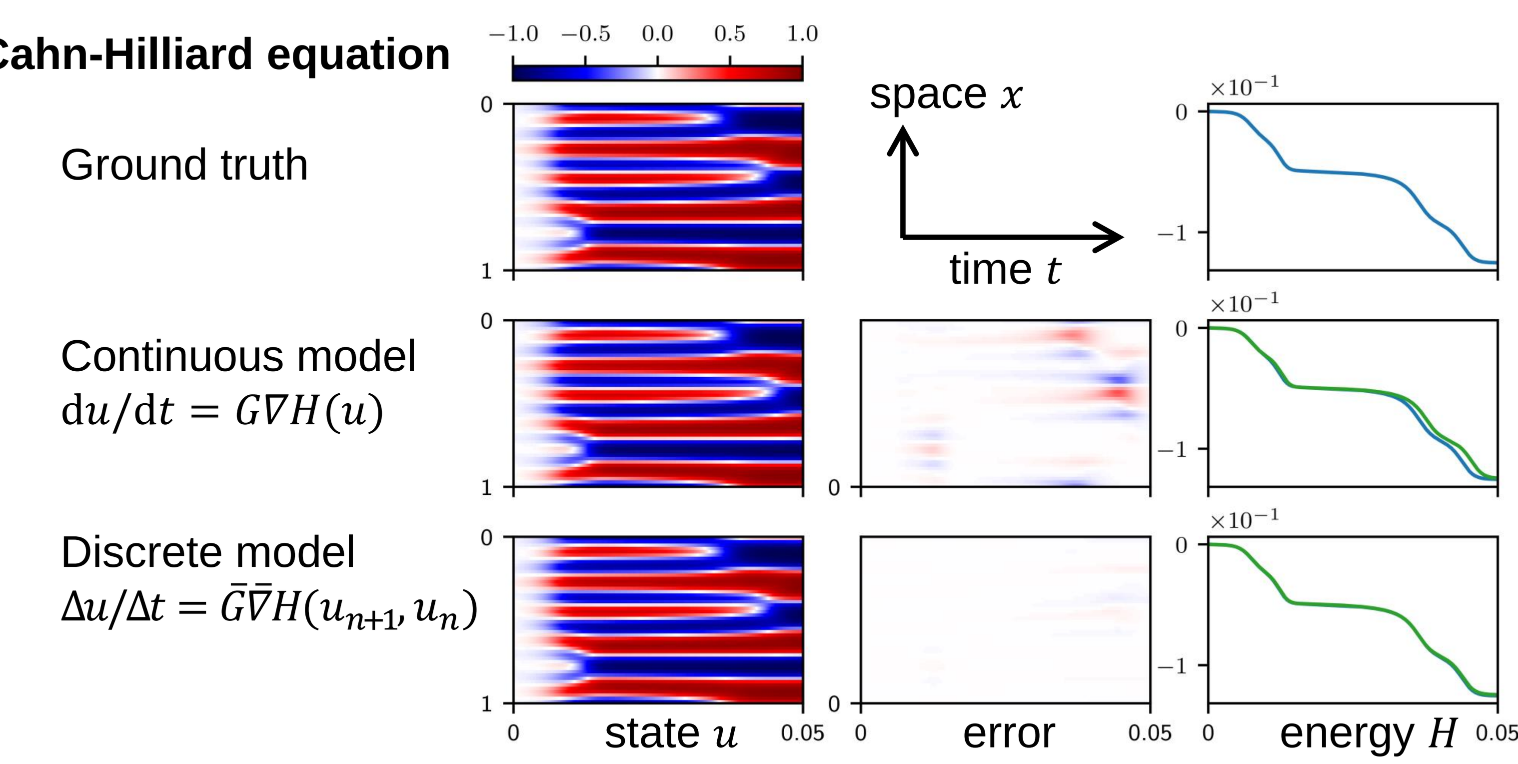
Models	KdV equation (conservative) energy H	KdV equation (conservative) state u	Cahn-Hilliard equation (dissipative) energy H	Cahn-Hilliard equation (dissipative) state u
Continuous model	3.01	0.34	4.89	0.80
Discrete model	**1.60**	**0.25**	**0.34**	**0.07**

KdV equation



The discrete model conserves the learned energy within a range of the rounding error, while the continuous model suffers from "energy drift".

Cahn-Hilliard equation



A numerical integrator tends to accumulate a modeling error over substeps, while the discrete model directly exhibits a discrete-time behavior.

ODE systems [1] (MSEs of the energy H in appropriate scales)

Models	conservative mass-spring	conservative pendulum	conservative 2-body	dissipative pendulum
Continuous model	1.74	16.55	81.84	3.44
Symplectic integrator	0.69	11.24	**40.37**	(9.64)
Discrete model	**0.62**	**10.79**	81.03	**0.50**

The symplectic integrator is specialized to a high-dimensional conservative system thanks to the separability assumption. A discrete model with such assumption is a future work.