Komatsuzaki Seminar
11:00~, March 9th, 2026.

Toward scientific discovery with AI
Part I

Yoh-ichi Mototake
Hitotsubashi university

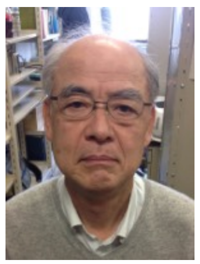
Self introduction

Born in Fukuoka City, Japan

Academic background:

4/2004 Faculty of Engineering (information science), Tohoku University (4/2004 ~ 9/2005)

3/2008 Faculty of Science, Tohoku University (10/2005 ~ 3/2008), **<u>Major in Physics</u>**



3/2010 Graduate School of Science, Hokkaido University
Master's Degree (Particle Physics, Theme: Wave packet dynamics), Supervisor: Prof. Kenzo Ishikawa

3/2013 Graduate School and College of Arts and Sciences, the University of Tokyo
Master's Degree (Cognitive Science, Theme: Human Agent Interaction), Supervisor: Prof. Kazuhiro Ueda

3/2016 Graduate School and College of Arts and Sciences, the University of Tokyo
Ph.D (**<u>Complex System</u>**, Theme: Deep Learning, Active matter) Supervisor: Prof. Takashi Ikegami

Work Experience:

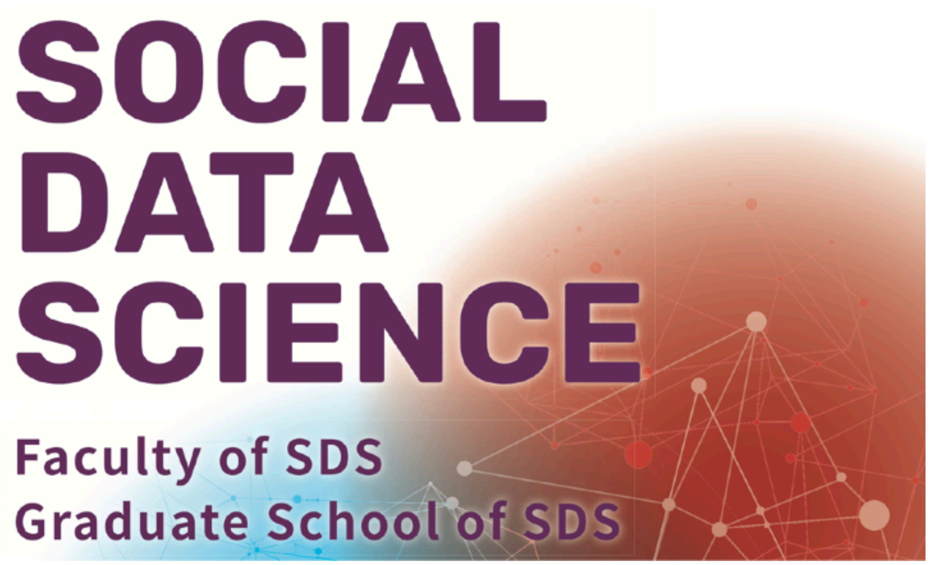
5/2016 〜 3/2019 Graduate School of Frontier Sciences, The University of Tokyo
Postdoctoral researcher (**<u>Data-driven science</u>**), worked at Prof. Masato Okada

4/2019 〜12/2022 The Institute of Statistical Mathematics, Research Center for Statistical Machine Learning
Projected assistant professor (**<u>Statistical Machine Learning</u>**), worked at Prof. Kenji Fukumizu

1/2023〜 Graduate school of Social Data Science, Hitotsubashi university
Associate Professor (**Data-driven science, Pattern dynamics, Interpretable AI**)

Our department

HITOTSUBASHI UNIVERSITY



Faculty Members

Social Science

NAOHIRO SHICHIJO
Science and Technology Policy, Computational Materials Science, Management of Technology

MAYU TERADA
Advanced Technology and Law, Administrative Law, Information Law

ATSUSHI HIYAMA
Human Augmentation, Virtual Reality, Gerontechnology

HIROTO KATSUMATA
Political Methodology, Political Representation

SUSUMU NAGAYAMA
Business Administration, Organization Theory, Creativity, Well-being

Statistics

SUSUMU IMAI
Labor Economics, Industrial Organization, Applied Econometrics

CHIHIRO SHIMIZU
Index Number and Theory, Applied Econometrics, Economic Measurement

TOSHIAKI WATANABE
Financial Econometrics, Macroeconometrics, Bayesian Econometrics

YOSHIMASA UEMATSU
Statistics, High-dimensional Data Analysis, Time Series Analysis

RYO KATO
Marketing Science, Bayesian Statistics, Missing Data Analysis

SHINICHIRO SHIROTA
Bayesian Statistics, Spatial/Spatio-temporal Statistics, Computational Statistics

Information Science

MAMORU KOMACHI
Computational Linguistics, Natural Language Processing, Artificial Intelligence

SHINSUKE SUZUKI
Neuroeconomics, Computational Neuroscience, Social Neuroscience, Computational Psychiatry

ATSUSHI KEYAKI
Information Retrieval, Natural Language Processing, Dialogue Systems

HARUAKI FUKUDA
Cognitive Science, Visual Perception, Cognitive Neuroscience

YOH-ICHI MOTOTAKE
Data Driven Science, Interpretable AI, Machine Learning

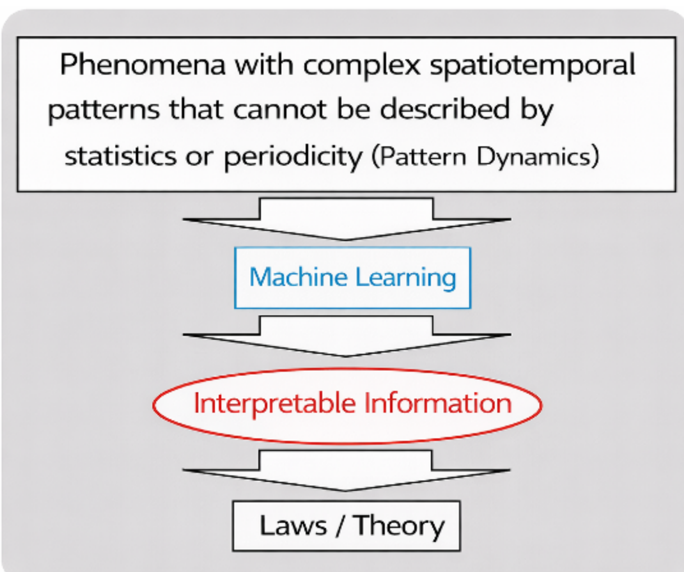
TATSUYA YATAGAWA
Computer Graphics, Geometry Processing, Image and Video Processing

Our laboratory

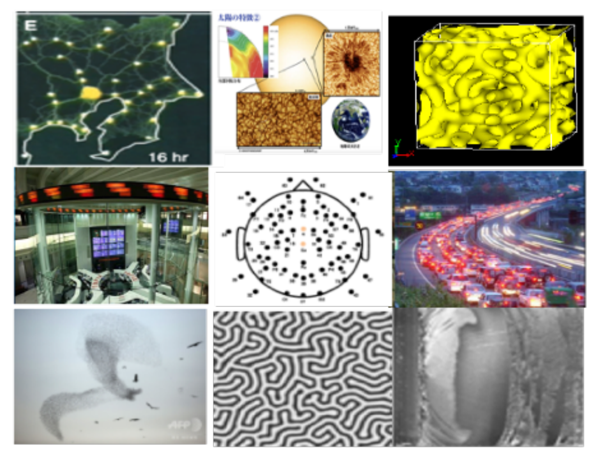
Data Driven Science Group

Mototake Lab.

Research Themes: Data-Driven Science, Interpretable AI, Machine Learning



【Target Phenomena】
Pattern dynamics across various fields



【Methodology Keywords】

Physics-informed ML, Bayesian modeling (Singular Bayesian theory), Topological data analysis

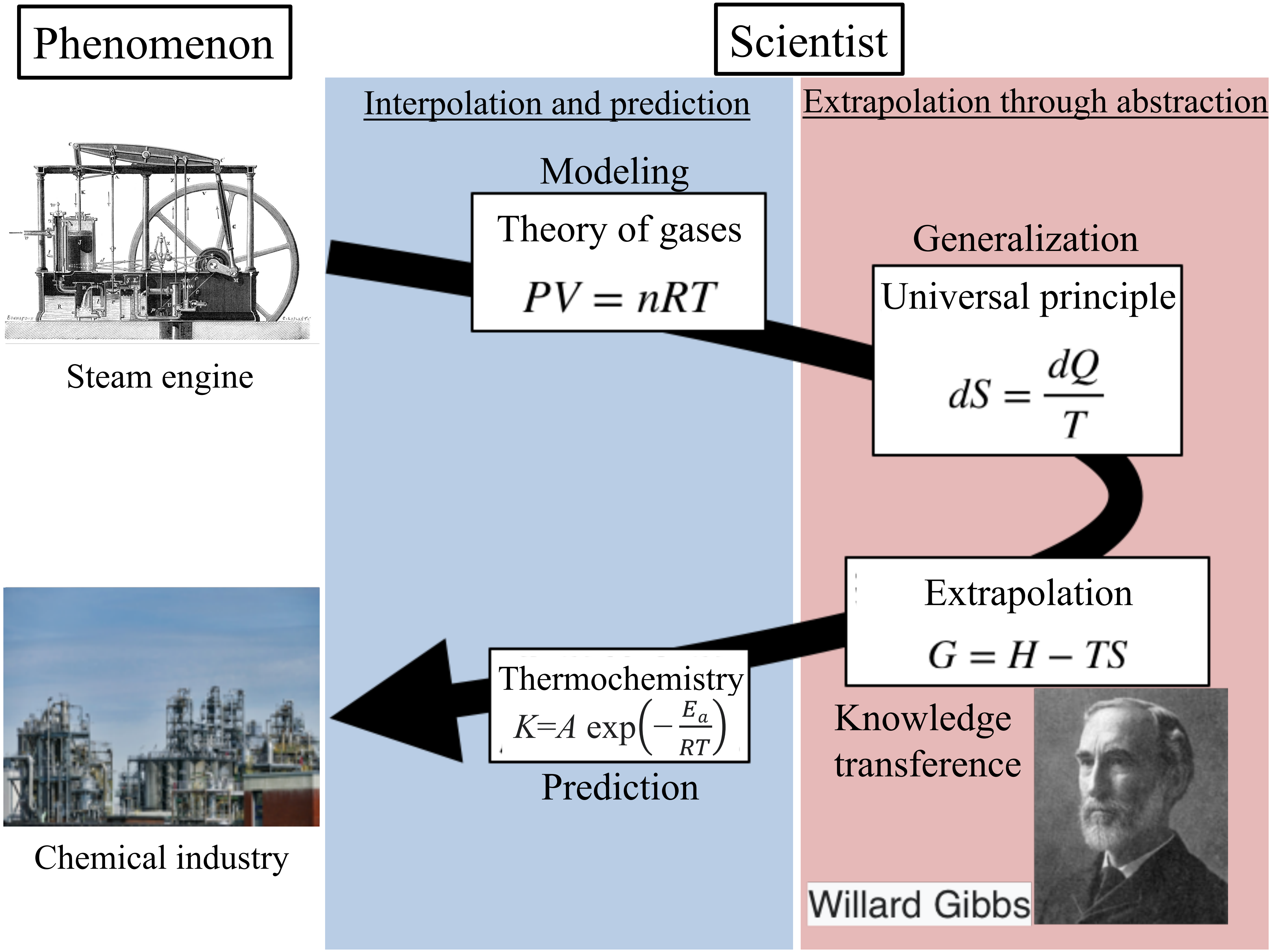
Laboratory organization (6 staff members, 6 students)

役職	名前	所属	専門領域
准教授（PI)	本武 陽一	一橋大学大学院 ソーシャル・データサイエンス研究科	物理学、解釈可能AI データ駆動科学
特任准教授	越智 久晃	一橋大学大学院 ソーシャル・データサイエンス研究科	MRIの物理 データ駆動科学
特任講師	森田 秀利	一橋大学大学院 ソーシャル・データサイエンス研究科	物理化学・数値計算 データ駆動科学

役職	名前	所属	専門領域
特任助教	幾田 佳	一橋大学大学院 ソーシャル・データサイエンス研究科	天文物理 データ駆動科学
特任助教	田中 賢	一橋大学大学院 ソーシャル・データサイエンス研究科	宇宙物理 データ駆動科学
講師（HIAS 連携教員)	有竹 俊光	一橋大学 社会科学高等研究院	機械学習 数理統計学

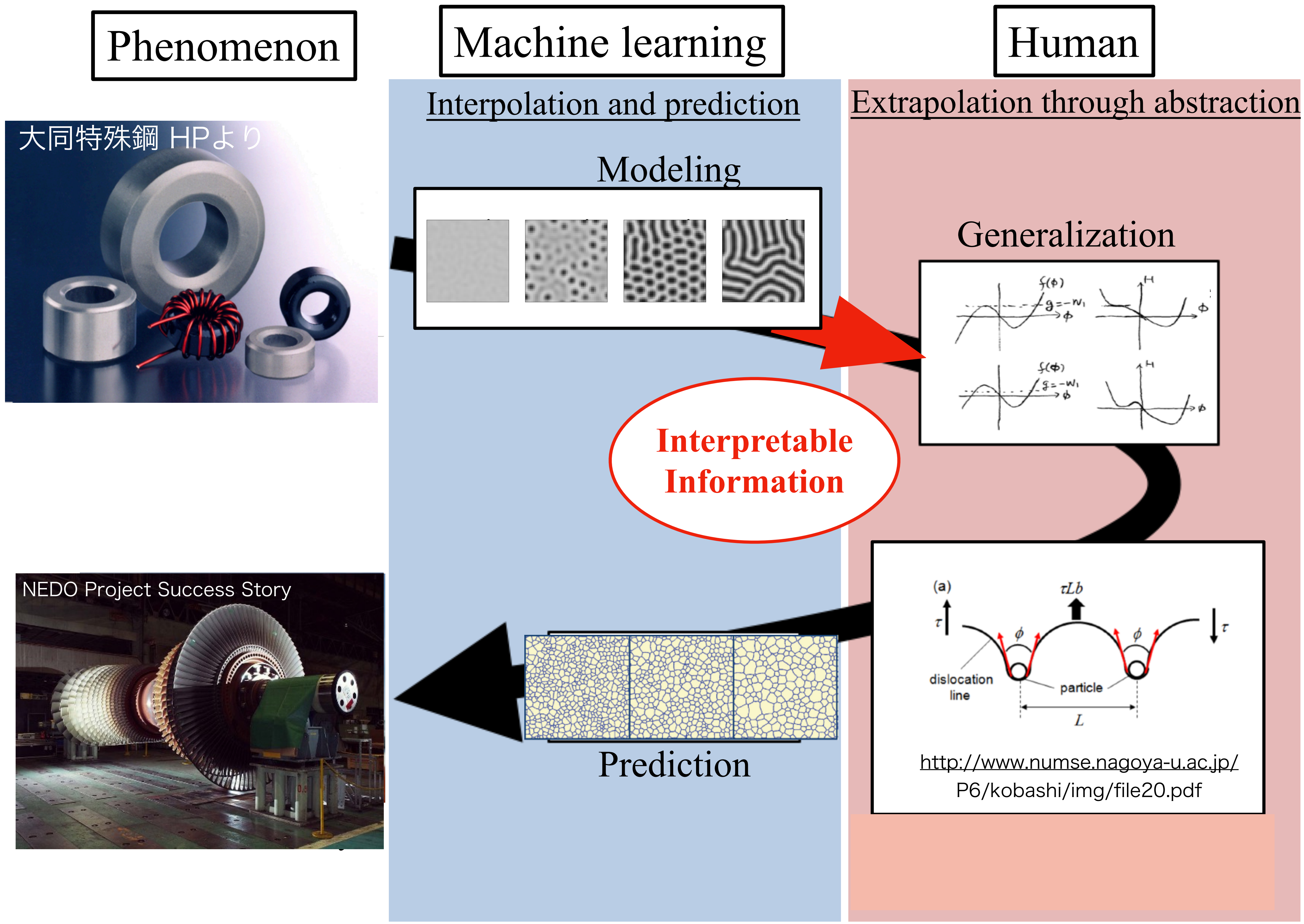
＋ 1 PhD, 3 Master's, 2 Undergraduate

Background: ML collaborating with Scientists



[鈴木炎,「エントロピーをめぐる冒険」, 2014]
[牧野功,「肥料製造技術の系統化」, 2008]
[武谷三段階論 [武谷 1942, 1966]]

Background: ML collaborating with Scientists



How to define interpretability?

Can interpretability be defined absolutely?

第236回研究会／第69回化合物新磁性材料専門研究会　https://www.magnetics.jp/event/topical_236/ ←Magnetic Materials Research Meeting

テーマ：

「新しい磁性研究のための量子ビームと計算科学の連携利用」

日　時:

2022年2月9日（水）13:00～18:00

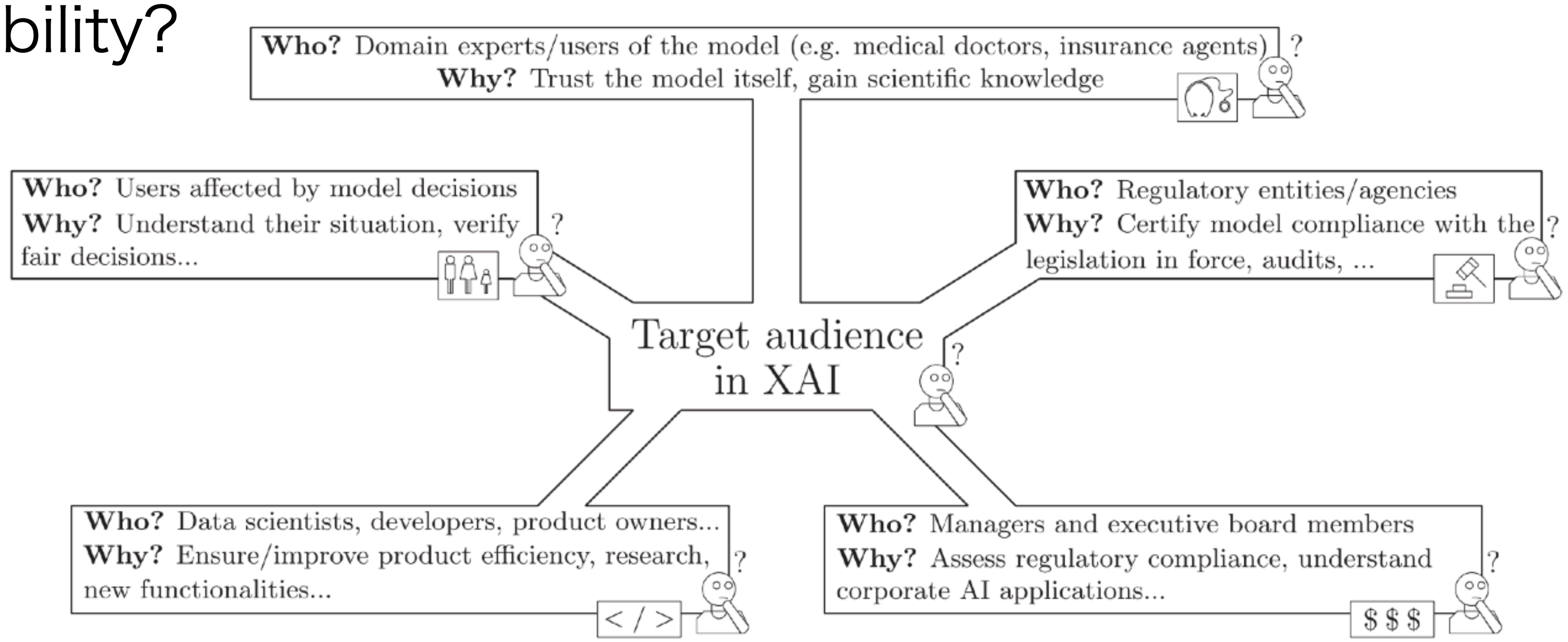
Since spectra have been observed by measurement since there were no computers, the "interpretation" of the structure of the Fourier spectrum may be the result of empirical and experiential learning of the correspondence between the structure and the spectrum in the original space.

takashi ikegami @alltbl · 5月15日

わかった、わかる、て心の動きは内蔵感覚の仕業であり、LLMでは作れない

→The mental process of "understanding" should be the work of the sense of internal organs and cannot be created by the LLM.

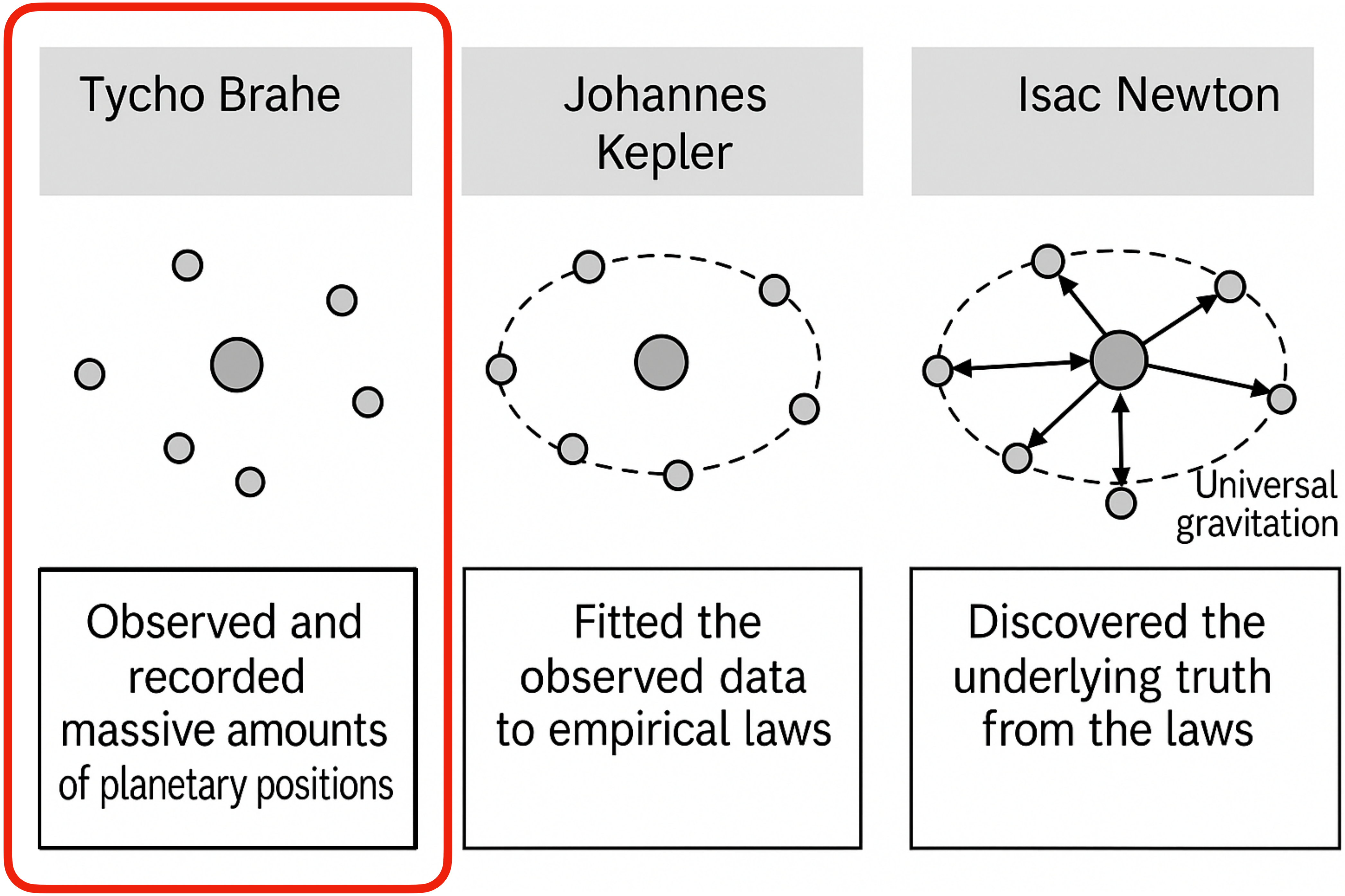
for whom is the interpretability?



[Arrieta, Alejandro Barredo, et al., Information fusion, 58, 2020.]

▶ **It is difficult to define "interpretability" in general. The first step is to observe the process to realize "interpretability".**

Like Tycho Brahe



[The figure was created based on ‘causal inference’ by Taku Kanemoto.]

Observing scientists' data-driven interpretation of complex phenomena to identify common problems and develop information science approaches to solve them.

Like Tycho Brahe

○ Observing the process of understanding using machine learning

My research project and collaborators

structural materials
井上純哉 教授（東大）
出村雅彦 部門長（NIMS）
森孝太郎 講師（茨城大）

Ceramic Composite Materials
佐藤光彦 教授（東京工科大）
久新荘一郎 教授（群馬大）

Nuclear fusion
仲田資季 准教授（核融合研）
森下侑哉 助教（京都大学）
小林進二 准教授（京都大学）
佐々木真 准教授（日本大学）

Polymer Materials
青柳岳司 センター長（旭化成）
青木裕之 教授（J-Parc/KEK）
天本義史 特任准教授（東京大）

Astrophysics
竹内努 准教授（名古屋大学）
政田洋平 准教授（福岡大学）

Bioinformatics
田口善弘 教授(中央大)

Information science
日野英逸 教授（統数研）
永田賢二 主任研究員（NIMS）
佐々木博昭 准教授
（はこだて未来大）

Applied mathematics
園田翔 研究員(理研)
大林一平 教授(岡山大)
有竹俊光　講師(一橋大)
奥野彰文 助教（統数研）

Nuclear Physics
三島賢二 準教授（KEK）
奥平琢也 助教（名古屋大）

Driving a car
自動車メーカー

Neutron beam measurement
二川 正敏 主任研究員（J-Parc）
涌井 隆 副主任研究員（J-Parc）

Condensed Matter Physics
水牧仁一朗 教授（熊大）
工藤和恵 准教授（お茶大/東北大）
赤井一郎 教授（熊大）
尾方成信 教授（大阪大）

Mechanics of Materials

Social Media Marketing
SNS企業

Dynamical system

Autonomous AI society

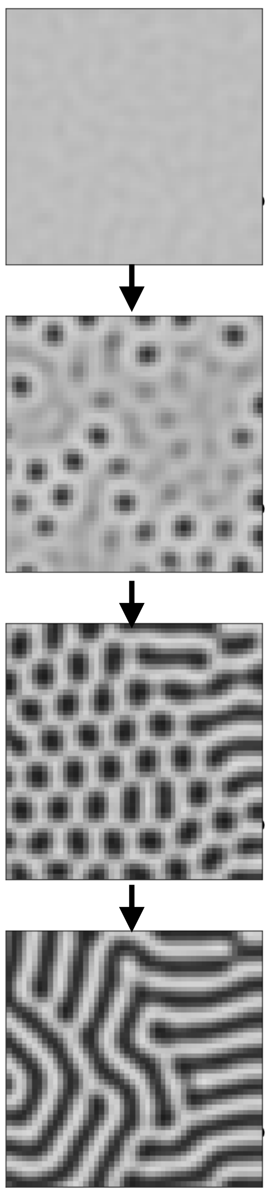
History

My approaches for Interpretable AI in physics

Our approach to interpretable AI

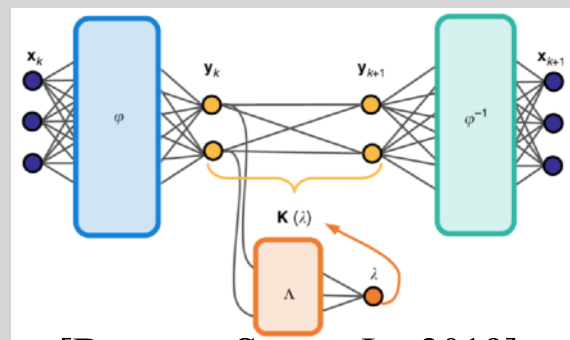
Complex phenomenon

dynamics



Part I **Part II**

Deep neural network-type approach



[Brunton, Steven L.. 2019]

Cons: low interpretability

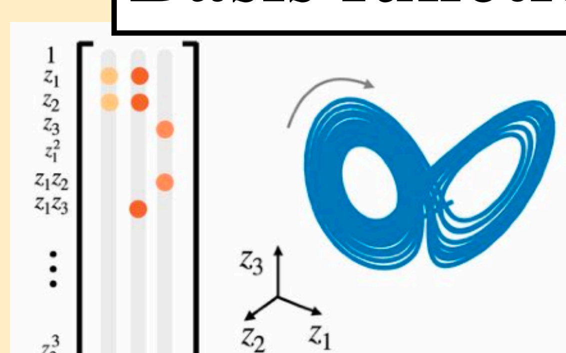
extract **Interpretable** physical information from DNN

Bayesian inference

Interpretable and high expressive model

Part III

Basis function type approach



[Steven L. Brunton et al., PNAS, 2016]

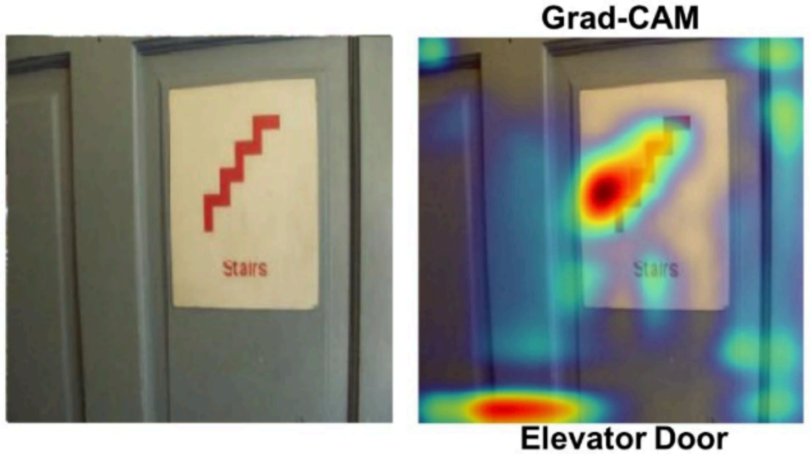
Cons: Poor expressive power

Construction of a reduced model based on **topological data analysis**

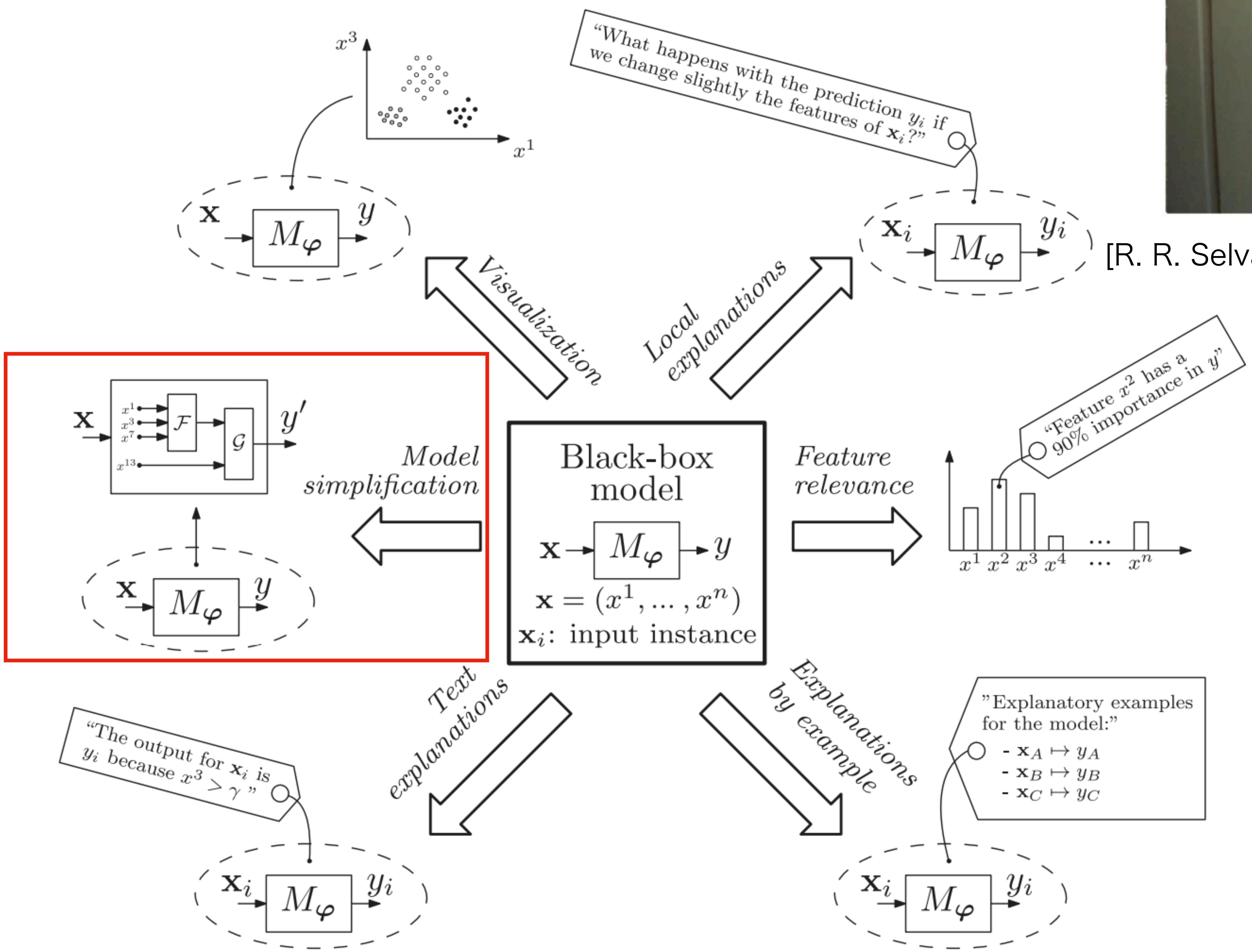
Expressive Power

Interpretability

Interpretable AI in DNN (Black box) type approach



[R. R. Selvaraju et al., IEEE ICCV, 618–626, 2017]

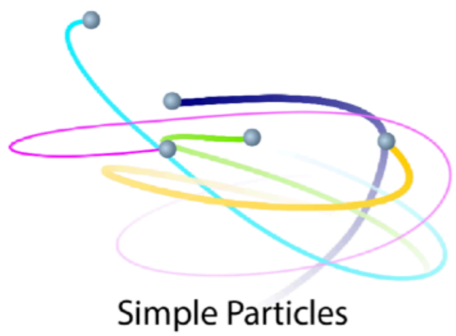


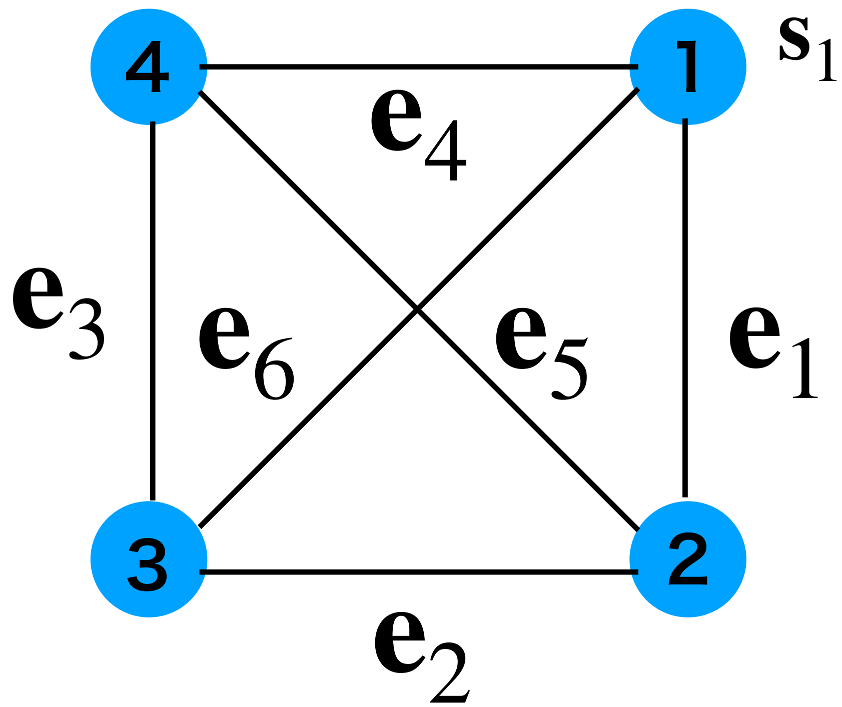
[Arrieta, Alejandro Barredo, et al., Information fusion, 58, 2020.]

▶ <u>In physics, do we want to ultimately express a physics law by combining elementary functions?</u>

○**Simplification of DNN by symbolic regression**

[Cranmer, Miles, et al., Neurips2020]

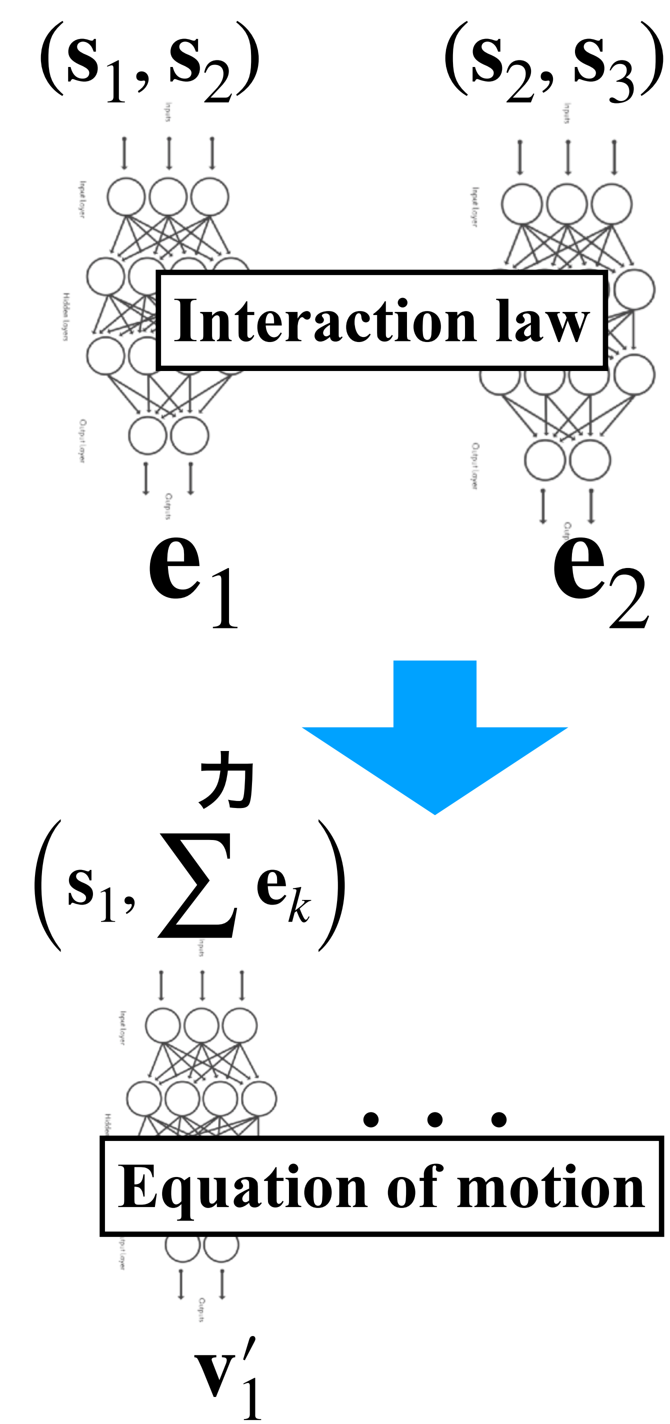




Interaction between 2 and 3

$\mathbf{s}_1 = (m_1, \mathbf{v}_1, \mathbf{r}_1, q_1, \ldots)$

State space of particle 1



Input graph $G = (V, E)$ with

nodes (e.g., positions of particles) $V = \{\mathbf{v}_i\}_{i=1:N^v};\ \mathbf{v}_i \in \mathbb{R}^{L^v}$, and

edges (indices of connected nodes) $E = \{(r_k, s_k)\}_{k=1:N^e};\ r_k, s_k \in \{1 : N^v\}$.

Compute messages for each edge: $\mathbf{e}_k' = \phi^e(\mathbf{v}_{r_k}, \mathbf{v}_{s_k})$,

$\mathbf{e}_k' \in \mathbb{R}^{L^{e'}}$, then

sum for each receiving node i : $\bar{\mathbf{e}}_i' = \sum_{k \in \{1:N^e | r_k = i\}} \mathbf{e}_k'$,

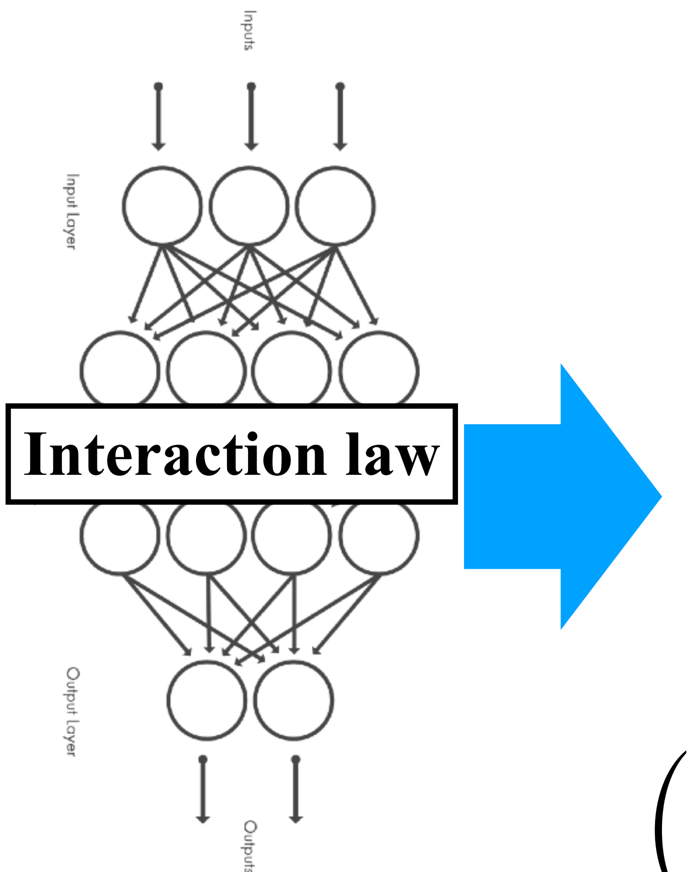
$\bar{\mathbf{e}}_i' \in \mathbb{R}^{L^{e'}}$.

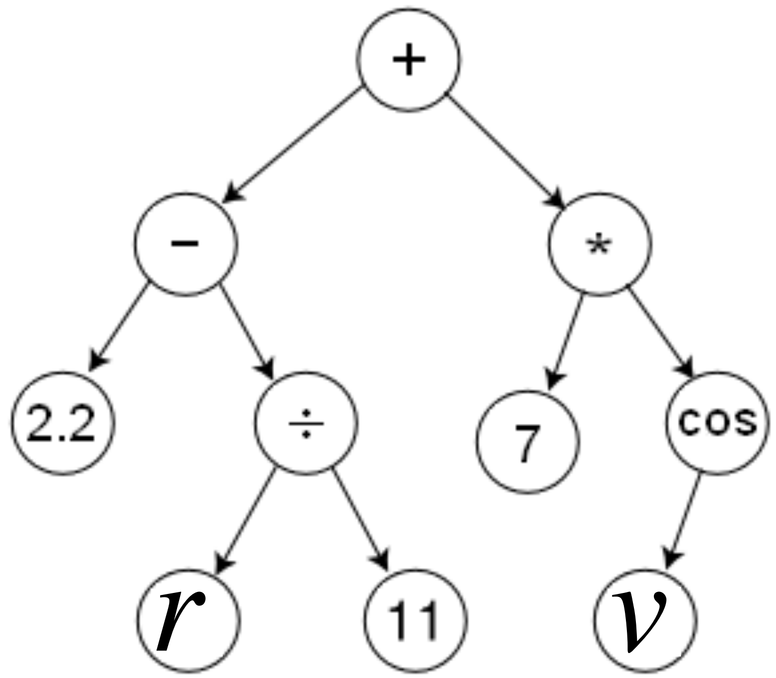
Compute output node prediction: $\hat{\mathbf{v}}_i' = \phi^v(\mathbf{v}_i, \bar{\mathbf{e}}_i')$

$\hat{\mathbf{v}}_i' \in \mathbb{R}^{L^{v'}}$.

Interpretable AI in DNN (Black box) type approach

○Simplification of DNN by symbolic regression





$$\left(2.2 - \left(\frac{r}{11}\right)\right) + (7.0 * \cos(v))$$

[https://en.wikipedia.org/wiki/Symbolic_regression]

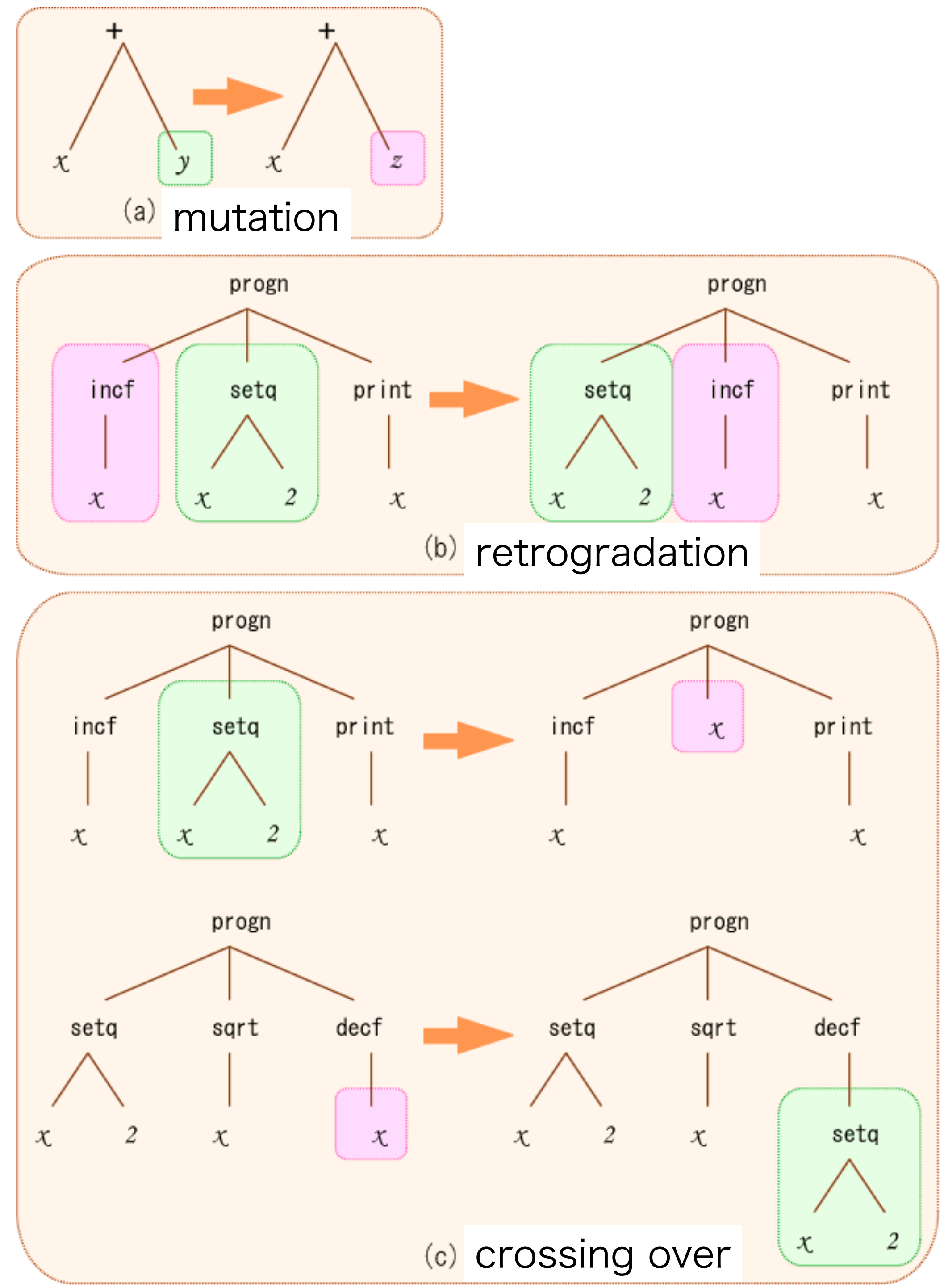
$$\text{Spring}: \; U_{12} = (r'_{12} - 1)^2$$

$$\text{Damped}: \; U_{12} = (r'_{12} - 1)^2 + \mathbf{r}_1 \cdot \dot{\mathbf{r}}_1 / n$$

$$\text{Charge}: \; U_{12} = q_1 q_2 / r'_{12}$$

$$\text{Dicontinuous}: \; U_{12} = \begin{cases} 0, & r'_{12} < 2 \\ (r'_{12} - 1)^2, & r'_{12} \geq 2 \end{cases}$$

genetic programming



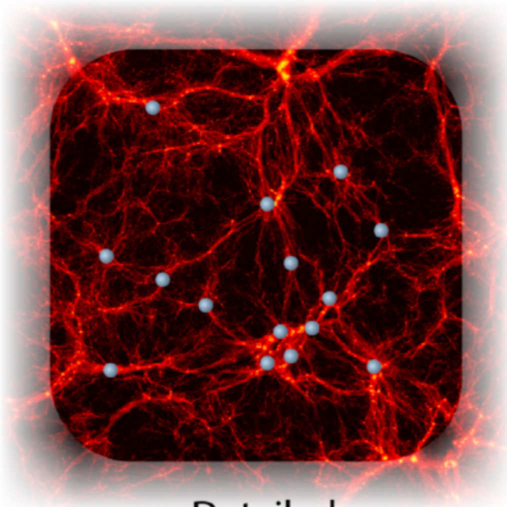
[from HP of Iba Lab.]

[Koza, John R., *Statistics and computing* 4.2 (1994): 87-112.]

Interpretable AI in DNN (Black box) type approach

○**Simplification of DNN by symbolic regression**

[Cranmer, Miles, et al., Neurips2020]



Detailed
Dark Matter Simulation

Consider dark matter halo as a particle, we want to know the law of the excess amount of matter around dark halo: $\delta = \frac{\rho - \langle\rho\rangle}{\langle\rho\rangle}$

	Test	Formula	Summed Component	$\langle\lvert\delta_i - \hat{\delta}_i\rvert\rangle$
Old	Constant	$\hat{\delta}_i = C_1$	N/A	0.421
Old	Simple	$\hat{\delta}_i = C_1 + (C_2 + M_i C_3) e_i$	$e_i = \sum_{j \neq i}^{\lvert \mathbf{r}_i - \mathbf{r}_j \rvert < 20} M_j$	0.121
New	Best, without mass	$\hat{\delta}_i = C_1 + \frac{e_i}{C_2 + C_3 e_i \lvert \mathbf{v}_i \rvert}$	$e_i = \sum_{j \neq i} \frac{C_4 + \lvert \mathbf{v}_i - \mathbf{v}_j \rvert}{C_5 + (C_6 \lvert \mathbf{r_i} - \mathbf{r_j} \rvert)^{C_7}}$	0.120
New	Best, with mass	$\hat{\delta}_i = C_1 + \frac{e_i}{C_2 + C_3 M_i}$	$e_i = \sum_{j \neq i} \frac{C_4 + M_j}{C_5 + (C_6 \lvert \mathbf{r}_i - \mathbf{r}_j \rvert)^{C_7}}$	0.0882

Table 2: A comparison of both known and discovered formulas for dark matter overdensity. C_i indicates fitted parameters, which are given in the appendix.

▶Discovered a novel law that better describes the phenomenon

Uncertainties in Physics-informed Inverse Problems: The Hidden Risk in Scientific AI

[Y. Mototake, M. Sasaki, **arXiv:2511.04564**]

[Y. Mototake, M. Sasaki, 15th Fusion Energy Alliance Meeting 2024]
[Y. Mototake, M. Sasaki, Japanese Joint Statistical Meeting 2024]

Yoh-ichi Mototake
Hitotsubashi university
Makoto Sasaki
Nihon university

Topics

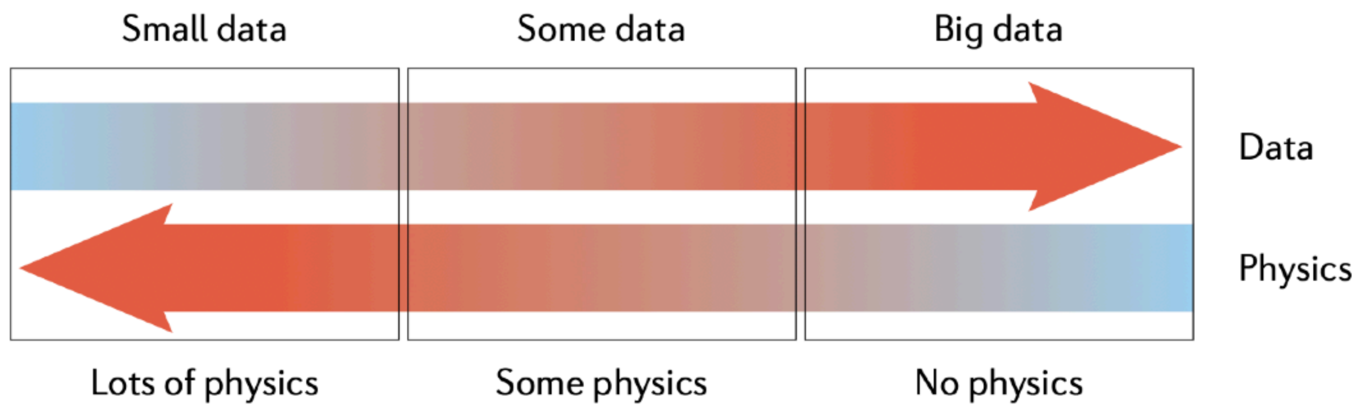
1. Physics-Informed Machine Learning and Scientific Discovery

2. Hidden Risks in Scientific Discovery via Physics-Informed Machine Learning

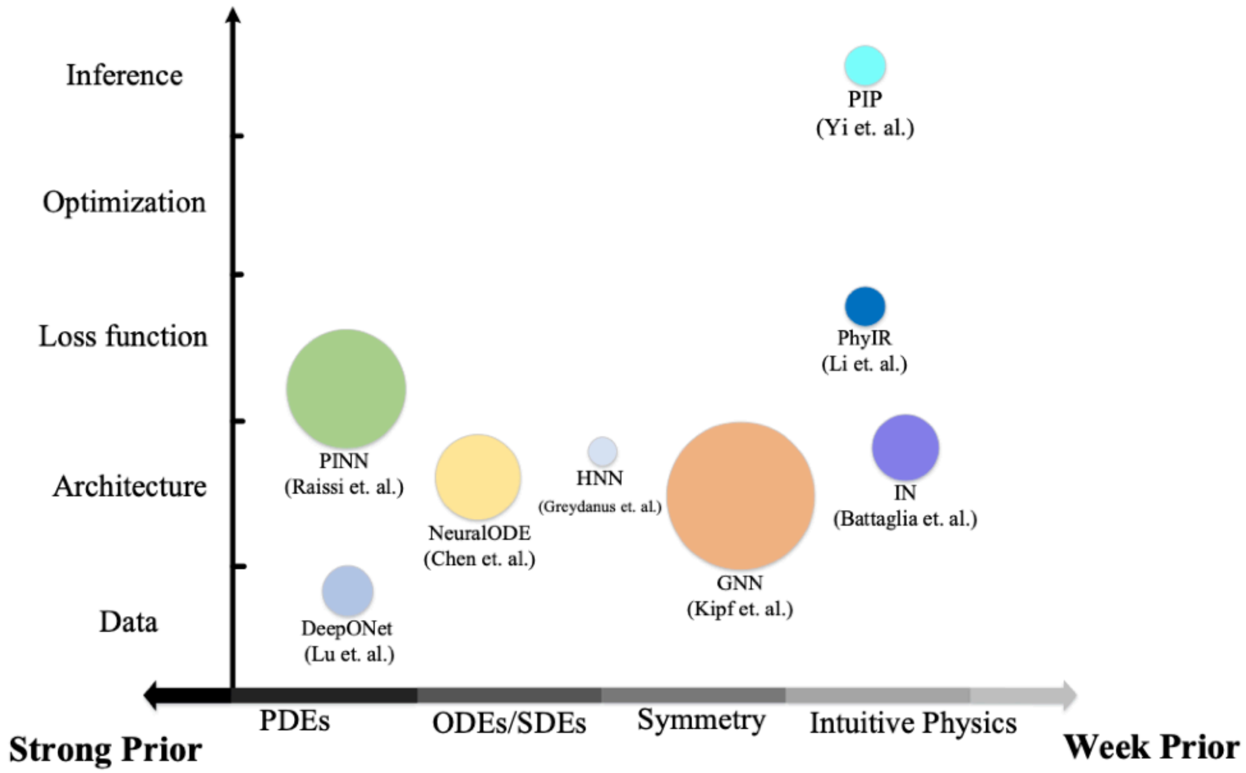
3. Examples of Risks and How to Deal with Them

○ Physics informed ML

→Physics informed ML is a framework that integrates **physical laws or governing equations** into **machine learning models.** [Karniadakis, G. E. et al., *Nature Reviews Physics*, *3*(6), 422-440.]



[Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L., *Nature Reviews Physics*, *3*(6), 422-440, (2021)]



(b)

[Hao, Z., Liu, S., Zhang, Y., Ying, C., Feng, Y., Su, H., & Zhu, J., *arXiv:2211.08064, 2022*]

Advantage:
It can achieve **high prediction accuracy** with a small amount of data.
It provides **higher interpretability** than modeling the entire data with a **black-box machine learning model**.

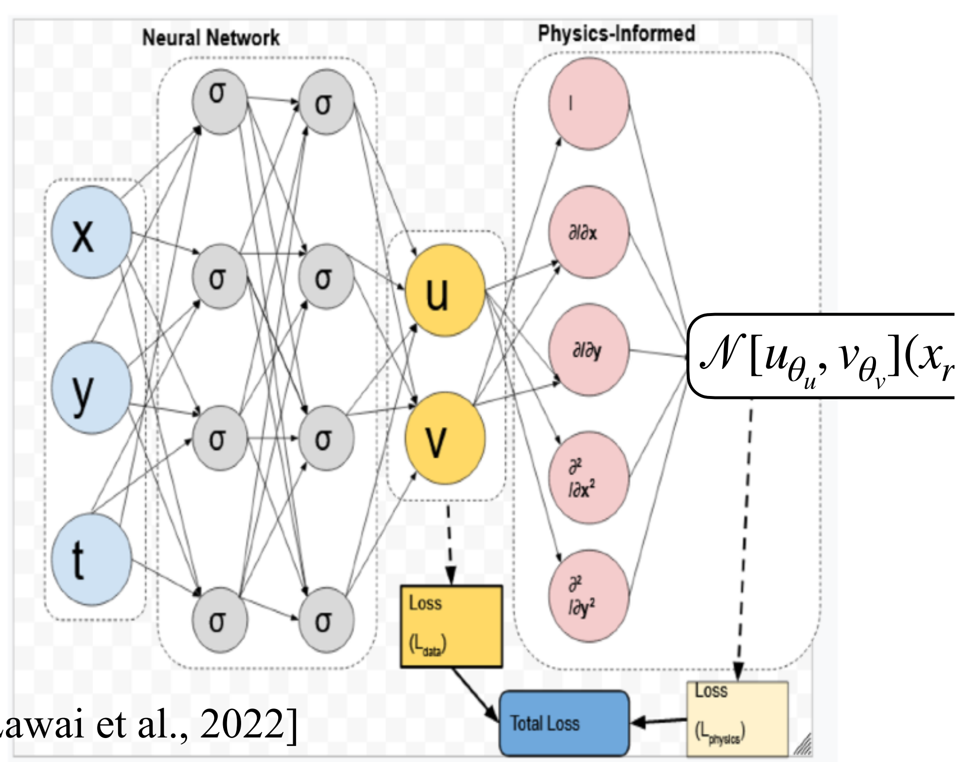
○ Physics informed ML

i.e. Physics-informed neural networks (PINN) [Raissi, M. et al.J. Comput. Phys., 2019.]
The governing PDE is typically of the form $\mathcal{N}[u, v](x) = 0, \quad x \in \mathbb{R}^d \times [0,T]$, where $\mathcal{N}$ is a nonlinear differential operator derived from physical laws, $u : \mathbb{R}^d \to \mathbb{R}$ is a sufficiently smooth function representing observation values, $v : \mathbb{R}^d \to \mathbb{R}$ is a sufficiently smooth coefficient function of PDE, and $x := (x, t) \in \mathbb{R}^d \times [0,T]$.

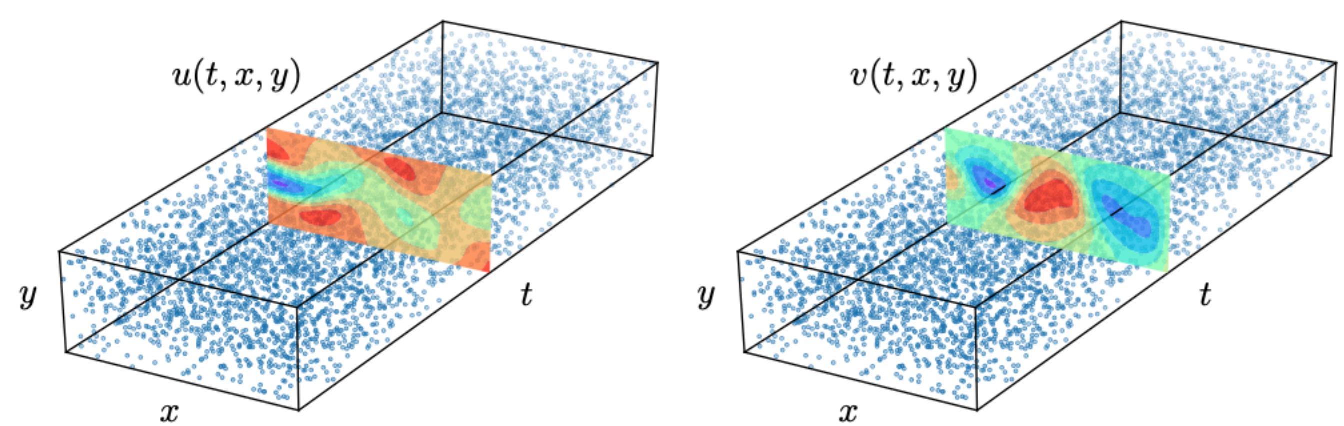
$$\mathcal{L}(\theta_u, \theta_v) = \mathcal{L}_{data}(\theta_u) + \lambda_{pde}\mathcal{L}_{pde}(\theta_u, \theta_v),$$

where $\mathcal{L}_{\mathrm{data}}(\theta_u) = \frac{1}{N}\sum_{i=1}^{N} \| u_\theta(x_i) - u_i \|^2$, $\mathcal{L}_{\mathrm{pde}}(\theta_u, \theta_v) = \frac{1}{N_r}\sum_{r=1}^{N_r} \| \mathcal{N}[u_{\theta_u}, v_{\theta_v}](x_r) \|^2$,

where $u_{\theta_u}(x)$ and $v_{\theta_v}(x)$ are neural network models parametrized by θ.

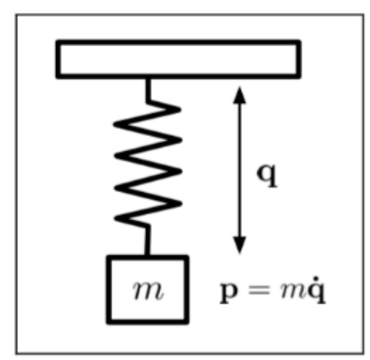


[Z. K. Lawai et al., 2022]



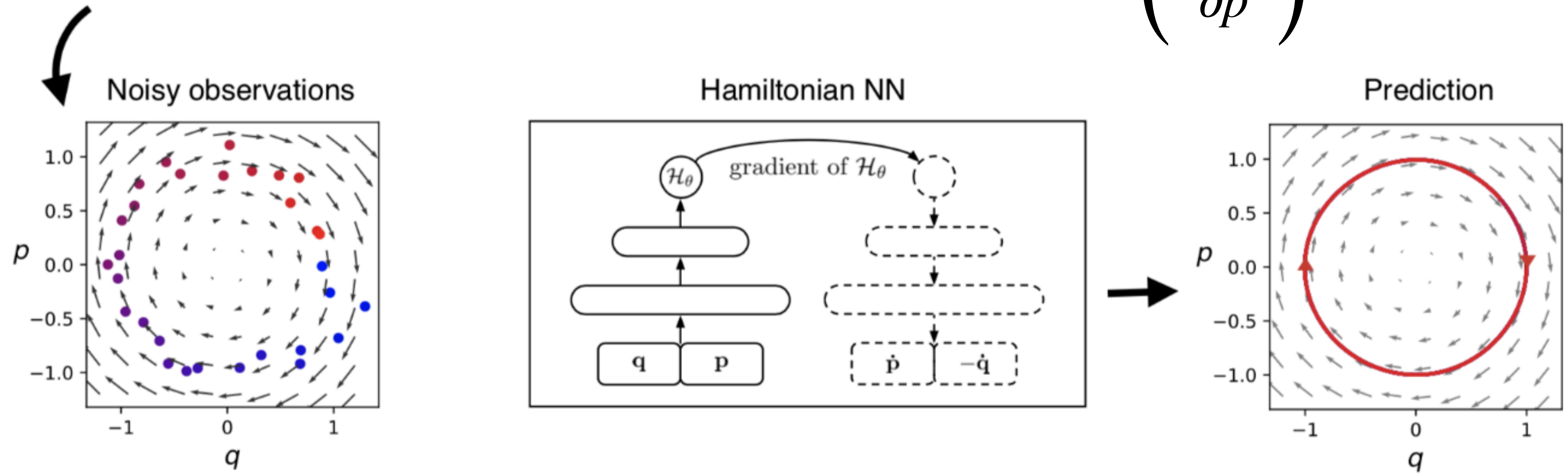
○ Physics informed ML

i.e. Hamiltonian neural networks [S. Greydanus, et al., NeurIPS, 2019]



↓Hamilton equation of motion with deep neural networks H_θ

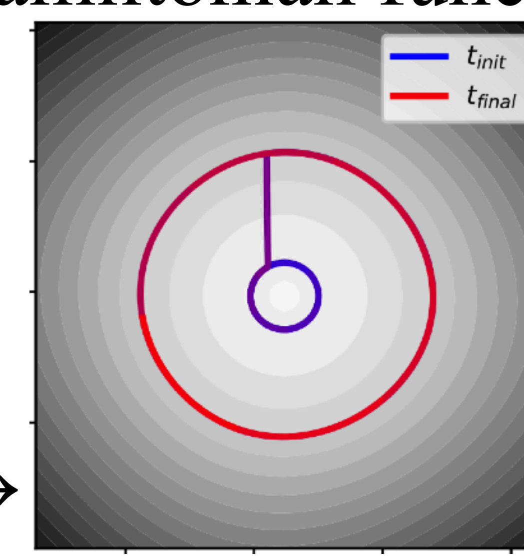
$$\begin{pmatrix} \dot{q} \\ \dot{p} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} \dfrac{\partial H_\theta}{\partial q} \\ \dfrac{\partial H_\theta}{\partial p} \end{pmatrix}$$



Train deep neural networks H_θ based on Hamilton equation of motion with

$$\mathcal{L}_{HNN} = \left\| \frac{\partial \mathcal{H}_\theta}{\partial \mathbf{p}} - \frac{\partial \mathbf{q}}{\partial t} \right\|_2 + \left\| \frac{\partial \mathcal{H}_\theta}{\partial \mathbf{q}} + \frac{\partial \mathbf{p}}{\partial t} \right\|_2$$

Hamiltonian function



Interpretability→

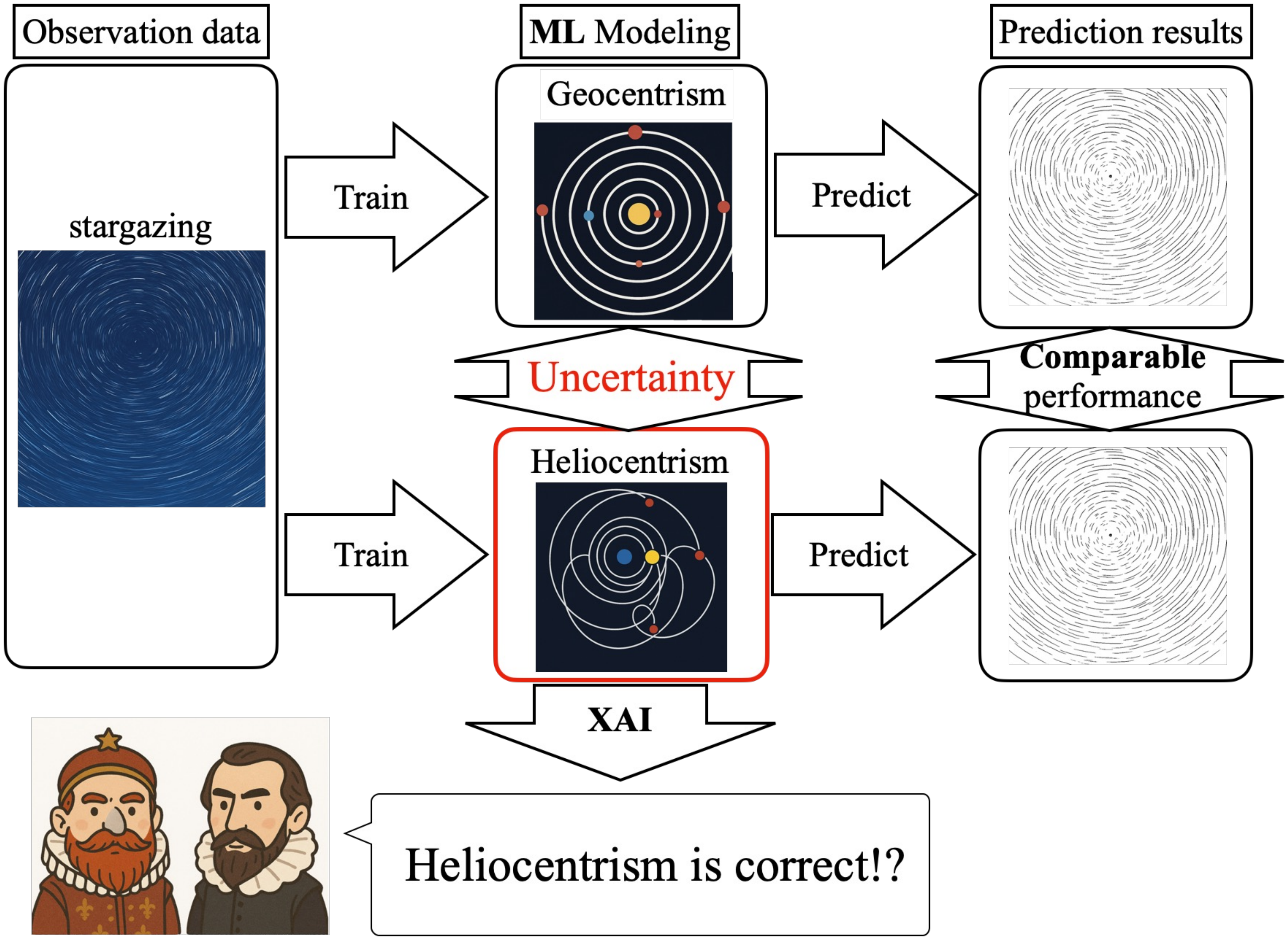
Topics

1. Physics-Informed Machine Learning and Scientific Discovery

2. Hidden Risks in Scientific Discovery via Physics-Informed Machine Learning

3. Examples of Risks and How to Deal with Them

The Hidden Risk in Physics Informed ML



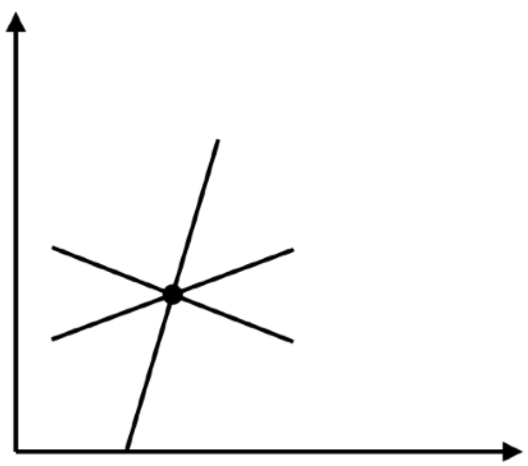
▶There are Hidden Risks in Physics informed ML

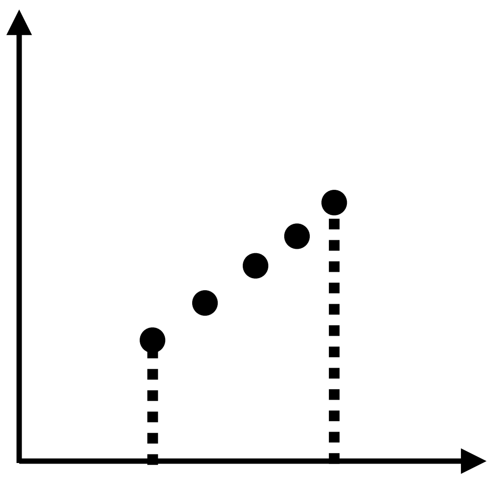
The Hidden Risk in Physics Informed ML

Three types of uncertainties

[Alexandre René and André Longtin., Nature Communications, 16(1):9393, 2025]
[Peter F. Pelz et. al., “Types of Uncertainty”, Springer International Publishing,, 2021]

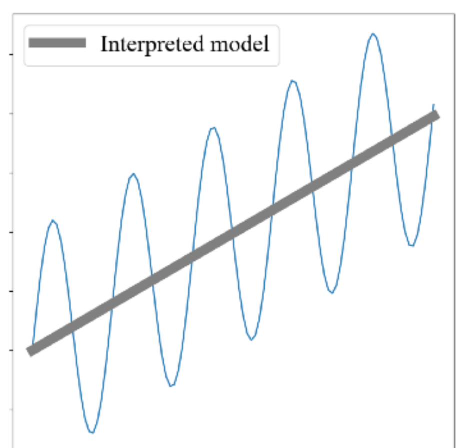
1. Data uncertainty





2. Model uncertainty

→Uncertainty due to model mismatch



3. Structural uncertainty

$$f(x; w_{\text{true}}) = f(x; w)$$

How to deal with

Increase the amount of data
Change how data is collected

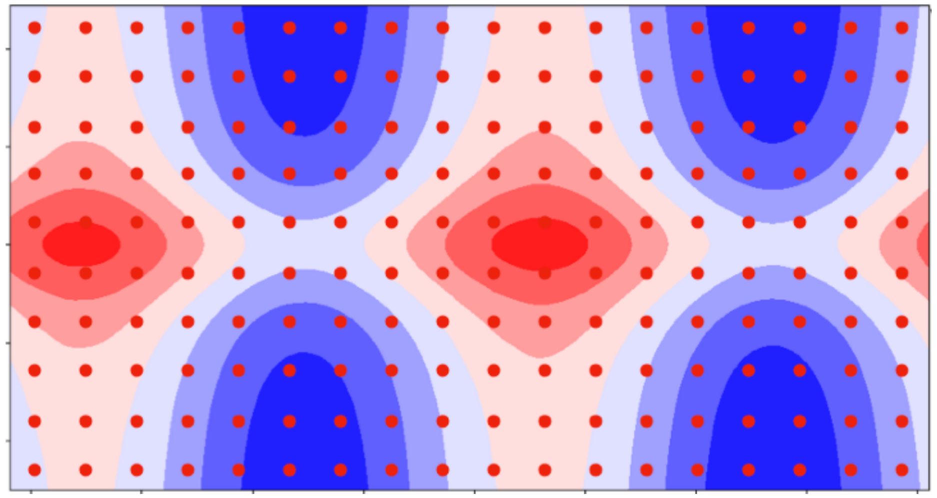
Improve the model

Add physical constraints

▶It is important to classify uncertainties and evaluate them in advance.

Definition [Structural Uncertainty under a PDE Constraint]

Let $a : \Omega \to \mathbb{R}$ denote the true underlying function, and let $u^\star$ be the corresponding state satisfying $\mathcal{N}(u^\star, a) = 0 \quad \text{on } \Omega\,.$ We say that **structural uncertainty** with respect to $a(x)$ arises if there exists another function $\hat{a}(x) \not\equiv a(x)$ such that $\mathcal{N}(u^\star, \hat{a}) = 0 \quad \text{on } \Omega\,.$



Theorem

Let consider m-th PDEs of the form

$$\sum_{\alpha \in A_{\geq m}} \varphi_\alpha^{(\ell)}(x, u(x), \partial u(x), \partial^2 u(x), \ldots) \cdot \partial^\alpha a(x) = C^{(\ell)},$$

where $C^{(\ell)}$ is constant, $\ell = 1,\ldots, L$, and $\varphi_\alpha^{(\ell)}$ is an arbitrary scalar function, $A_{\geq m} := \{\alpha \mid m \leq |\alpha|\}$, $m \geq k$, and $\partial^k u(x)$ represents the arbitrary set of k-th-order partial differential coefficients. That is, PDEs are linear in $\partial^\alpha a(x)$.

Assume that PDEs have a k-leaf set S_k^{leaf} in their equations. Discretize Ω on an infinitesimal grid with spacing $\varepsilon > 0$, and denote the grid points as $x^{(i)} \in \mathbb{R}^d$ for $i = 1,\ldots, N$. For each grid point, consider the discretized PDE system:

$$\sum_{\alpha \in A_{\geq m}} \varphi_\alpha^{(\ell)} \left(x_{(i)}, u(x_{(i)}), \partial u(x_{(i)}), \partial^2 u(x_{(i)}), \ldots \right) \partial^\alpha a(x_{(i)}) = C_{(i)}^{(\ell)}, \quad \ell = 1,\ldots, L.$$

Then, by stacking the equations across all grid points, we can represent the system as a linear system: $\mathbf{M} \cdot \mathbf{a} = \mathbf{c}$, where $\mathbf{a} \in \mathbb{R}^{|A_{\geq m}| \cdot N}$ is the vector of derivatives of the coefficient function $a(x)$, $\mathbf{M} \in \mathbb{R}^{LN \times |A_{\geq m}| N}$ is the matrix constructed from $\varphi_\alpha^{(\ell)}$, and $\mathbf{c} \in \mathbb{R}^{|A_{\geq m}| \cdot N}$ is the vector of constant values.

Then, the following statements are equivalent:

(i) $\mathrm{rank}(\mathbf{M}) = \mathrm{rank}([\mathbf{M}\ \mathbf{c}]) = |A_{\geq m}| N$

(ii) The k-th derivatives of coefficient function **a** has no structural uncertainty.

i.e. Hamiltonian system

Let the canonical variables be denoted by $(\boldsymbol{q}, \boldsymbol{p})^\top \in \mathbb{R}^{2d}$, where $\boldsymbol{q} = (q_1, \ldots, q_d)^\top$ are the generalized coordinates and $\boldsymbol{p} = (p_1, \ldots, p_d)^\top$ are the generalized momenta. Given the Hamiltonian function $H(\boldsymbol{q}, \boldsymbol{p})$, the canonical equations of motion (Hamilton's equations) can be expressed in matrix form as

$$\begin{pmatrix} \frac{d\boldsymbol{q}}{dt} \\ \frac{d\boldsymbol{p}}{dt} \end{pmatrix} =: \begin{pmatrix} \dot{\boldsymbol{q}} \\ \dot{\boldsymbol{p}} \end{pmatrix} = \begin{pmatrix} \mathbf{0} & I_n \\ -I_n & \mathbf{0} \end{pmatrix} \begin{pmatrix} \frac{\partial H}{\partial \boldsymbol{q}} \\ \frac{\partial H}{\partial \boldsymbol{p}} \end{pmatrix}.$$

The equation on an infinitesimal N grid space is written as

$$\mathbf{M} \cdot \mathbf{a} = \mathbf{c}\$, where\$M = \begin{pmatrix} \mathbf{0} & I_{Nd} \\ -I_{Nd} & \mathbf{0} \end{pmatrix},$$

$\mathbf{a} = \left(\frac{\partial H}{\partial \boldsymbol{q}_1}, \ldots, \frac{\partial H}{\partial \boldsymbol{q}_N}, \frac{\partial H}{\partial \boldsymbol{p}_1}, \ldots, \frac{\partial H}{\partial \boldsymbol{p}_N} \right)^\top \$, and\$ \mathbf{c} = \left(\dot{\boldsymbol{q}}_1, \ldots, \dot{\boldsymbol{q}}_N, \dot{\boldsymbol{p}}_1, \ldots, \dot{\boldsymbol{p}}_N \right)^\top$. Since $\mathrm{rank}(\boldsymbol{M}) = 2Nd$, the necessary conditions are satisfied such that the Hamiltonian function $H(\boldsymbol{q}, \boldsymbol{p})$ is uniquely determined, except for the indefiniteness of the constant.

i. e. Lagrange system

Let the generalized coordinates be denoted by $\boldsymbol{q} = (q_1, \ldots, q_d)^\top$.
Lagrange's equations of motion can be written in matrix form as

$$\dot{\boldsymbol{p}} := \frac{d}{dt}\frac{\partial L}{\partial \dot{\boldsymbol{q}}} = \begin{pmatrix} \mathbf{0} & I_d \end{pmatrix} \begin{pmatrix} \frac{\partial L}{\partial \boldsymbol{q}} \\ \frac{\partial L}{\partial \dot{\boldsymbol{q}}} \end{pmatrix}.$$

The equation on an infinitesimal N grid space is written as $\mathbf{M} \cdot \mathbf{a} = \mathbf{c}$, where $\mathbf{M} = \begin{pmatrix} \mathbf{0} & I_{Nd} \end{pmatrix}$,
$\mathbf{a} = \left(\frac{\partial L}{\partial \boldsymbol{q}_1}, \ldots, \frac{\partial L}{\partial \boldsymbol{q}_N}, \frac{\partial L}{\partial \dot{\boldsymbol{q}}_1}, \ldots, \frac{\partial L}{\partial \dot{\boldsymbol{q}}_N}, \right)^\top$\$, and $\mathbf{c} = \left(\dot{\boldsymbol{p}}_1, \ldots, \dot{\boldsymbol{p}}_N\right)^\top$.
Since $\text{rank}(\boldsymbol{M}) = Nd < 2Nd$, the Lagrange function $L(\boldsymbol{q}, \dot{\boldsymbol{q}})$ is undetermined.
The reason the Lagrangian cannot be determined is that the information corresponding to the part of the canonical equation of motion in the Hamiltonian system, $\dot{\boldsymbol{q}} := \frac{\partial H}{\partial \boldsymbol{p}}$, is missing in the Lagrange system.
Since one physical constraint for estimating the coefficient function has disappeared, the Lagrange function is not determined.
This missing information corresponds to the definition of the generalized momentum in the Lagrange system, $\boldsymbol{p} := \frac{\partial L}{\partial \dot{\boldsymbol{q}}}$.
In fact, adding the definition of generalized momentum to the Lagrangian equation of motion leads to the satisfaction of the necessary condition, $\text{rank}(\boldsymbol{M}) = 2Nd$, for the Lagrangian to be uniquely determined.

Three step framework to The Hidden Risk in Scientific AI

Step 1 (Diagnose): estimate rank/deficiency of M before ML training.

Step 2 (Constrain): add physics (symmetry, gauge fixing, boundary info) to shrink ker(M).

Step 3 (Sweep): vary constraint strength λ and inspect solution families (don't tune by prediction loss only).

Topics

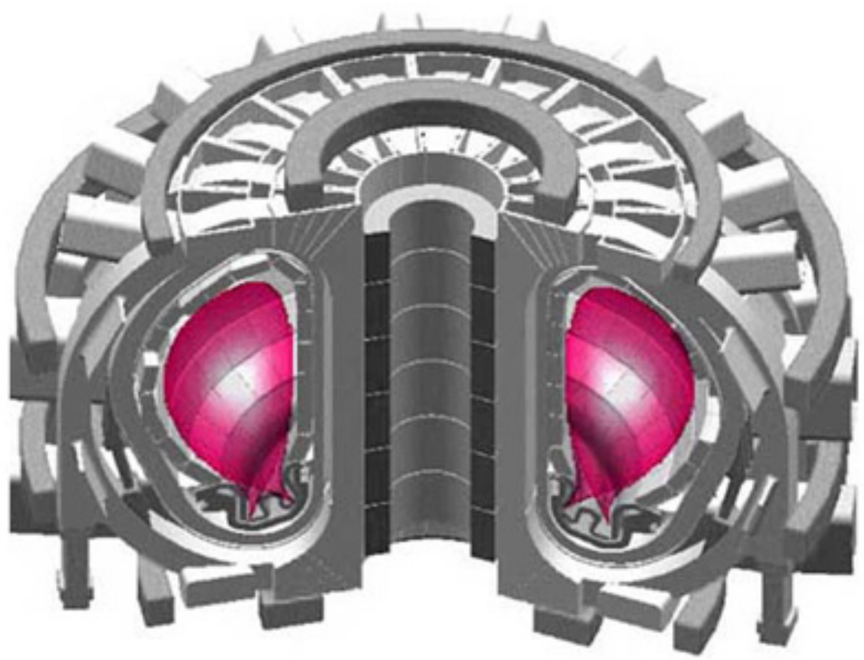
1. Physics-Informed Machine Learning and Scientific Discovery

2. Hidden Risks in Scientific Discovery via Physics-Informed Machine Learning

3. Examples of Risks and How to Deal with Them

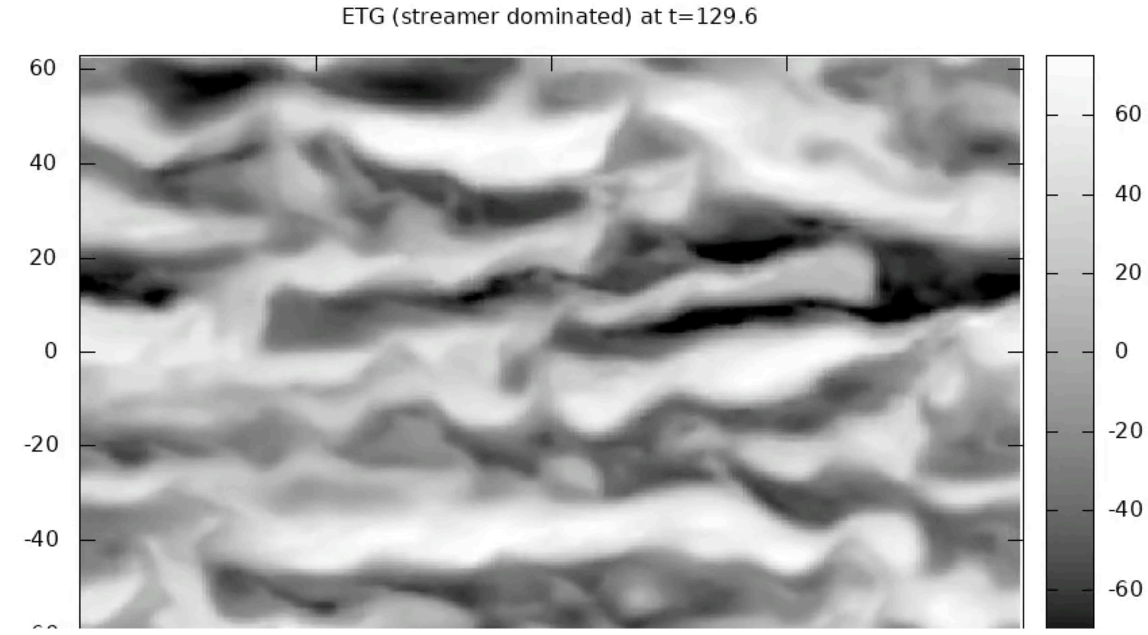
Zonal flow system in nuclear reactor

○ **Plasma turbulence interacting with Zonal flow can maintain a plasma state:**

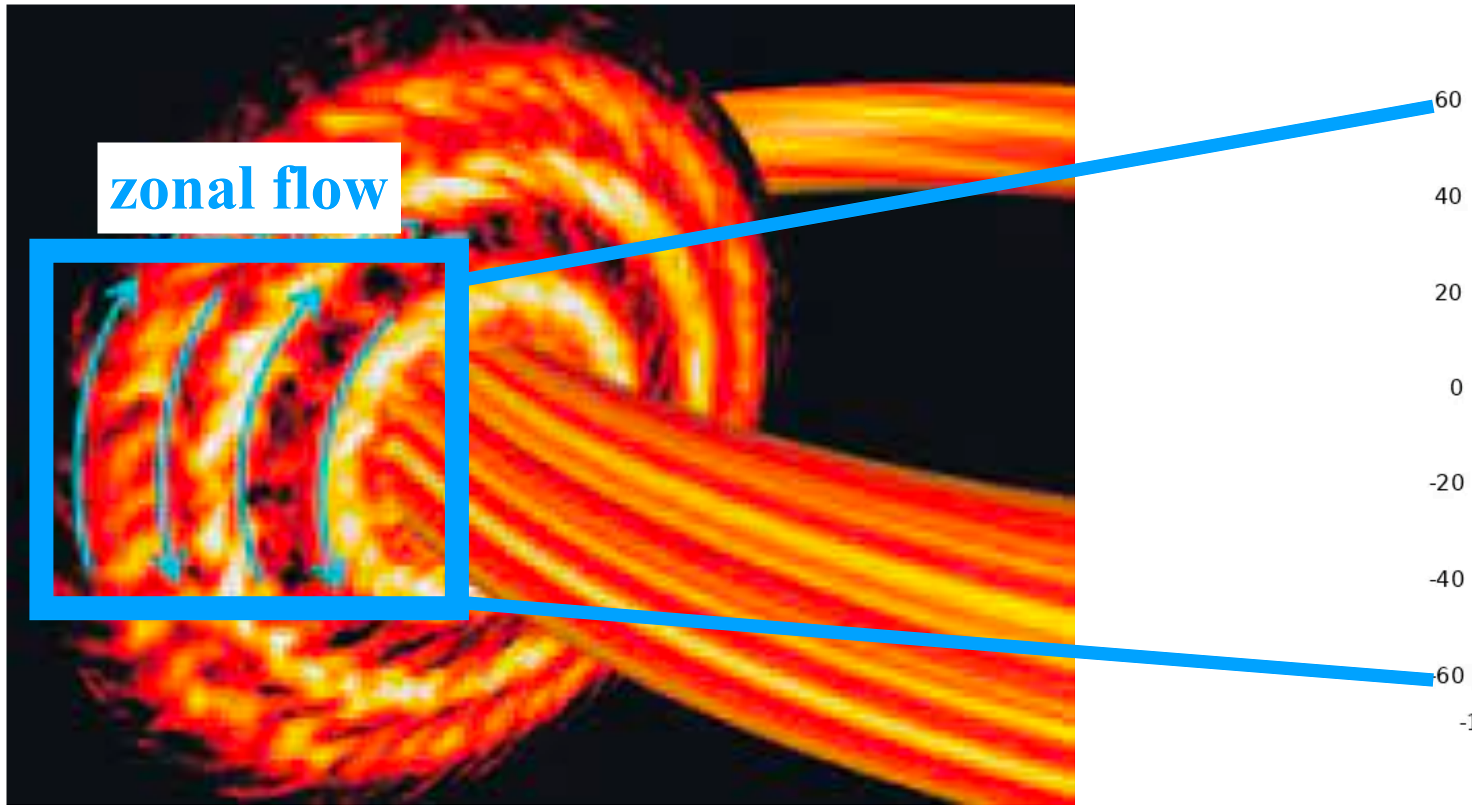


https://www.fusion.qst.go.jp/ITER/iter/page1_7.html

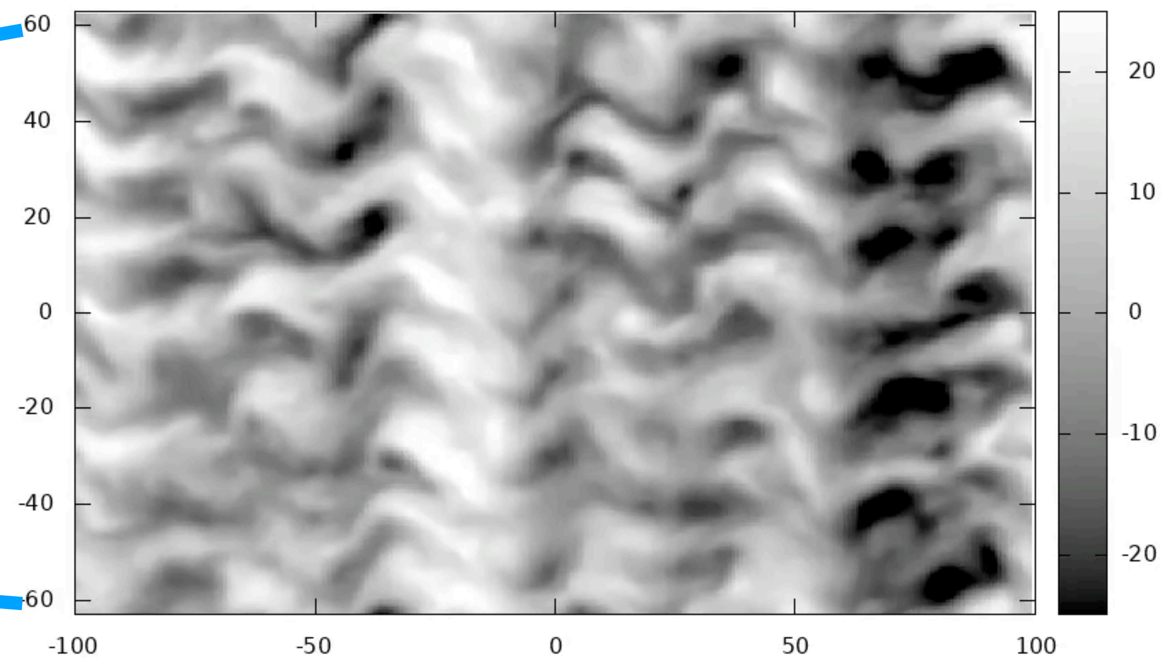
Heat transport by turbulence



[provided by Prof. Nakata NIFS]



Zonal flow



[provided by Prof. Nakata NIFS]

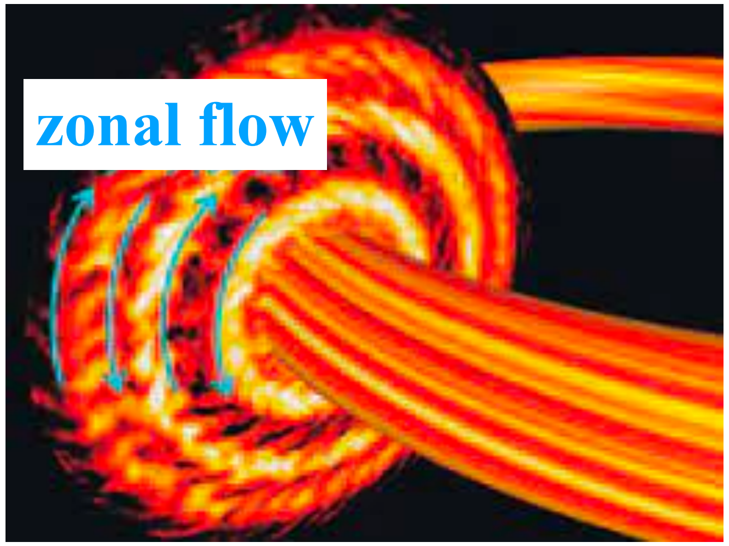
Reduced model of zonal flow system

○ **Wave kinetic theory of vortex in plasma turbulence interacting with Zonal flow**

[A. I. Smolyakov; P. H. Diamond, Phys. of Plasmas, 1999]
[A. I. Smolyakov, P. H. Diamond, and V. I. Shevchenko, Phys. of Plasmas, 2000]

Hamiltonian of vortexes interacting with Zonal flow $V_y(x,t)$:

$$H(x, k_x, t) = H_0 + \frac{k_y}{1 + k_x^2 + k_y^2} + k_y V_y(x, t)$$



Time evolution equation of zonal flow :

$$\frac{\partial V_y(q,t)}{\partial t} = \frac{\partial^2}{\partial q^2} \int dp \frac{p k_y I(q,p,t)}{\left(1 + p^2 + k_y^2\right)^2} + \mu \frac{\partial^2 V_y(q,t)}{\partial q^2}$$

Time evolution of intensity of vortex $I(q,p)$:

$$\frac{\partial I(x,k_x,t)}{\partial t} + \frac{\partial H(x,k_x,t)}{\partial k_x}\frac{\partial I(x,k_x,t)}{\partial x} - \frac{\partial H(x,k_x,t)}{\partial x}\frac{\partial I(x,k_x,t)}{\partial k_x} = C,$$

$C := \gamma_L(p) I(x,k_x,t) - \Delta\omega [I(x,k_x,t)]^2$: linear growth + growth restriction of turbulence:

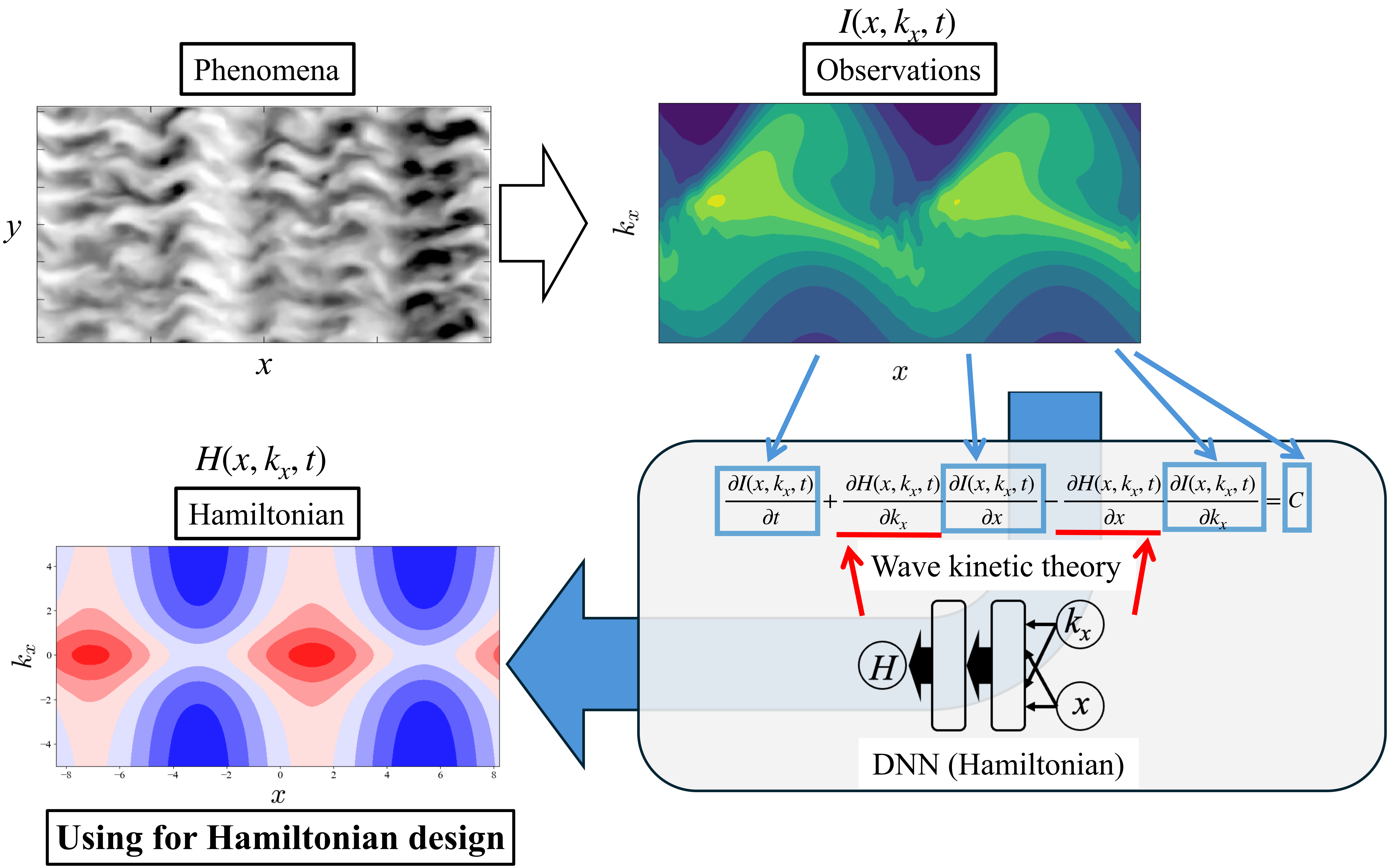
$$\gamma_L(k_x) = \frac{k_y(p^2 + k_y^2)}{D(1 + p^2 + k_y^2)^3} \exp\left(-\left(\frac{k_y}{\Delta k}\right)^2\right), D, \Delta k = \text{const}.$$

Δw : Width of spectrum of $I(x, k_x)$

▶It is difficult to design a Hamiltonian that precisely fits the experimental data

○ **Wave kinetic theory of vortex in plasma turbulence interacting with Zonal flow**

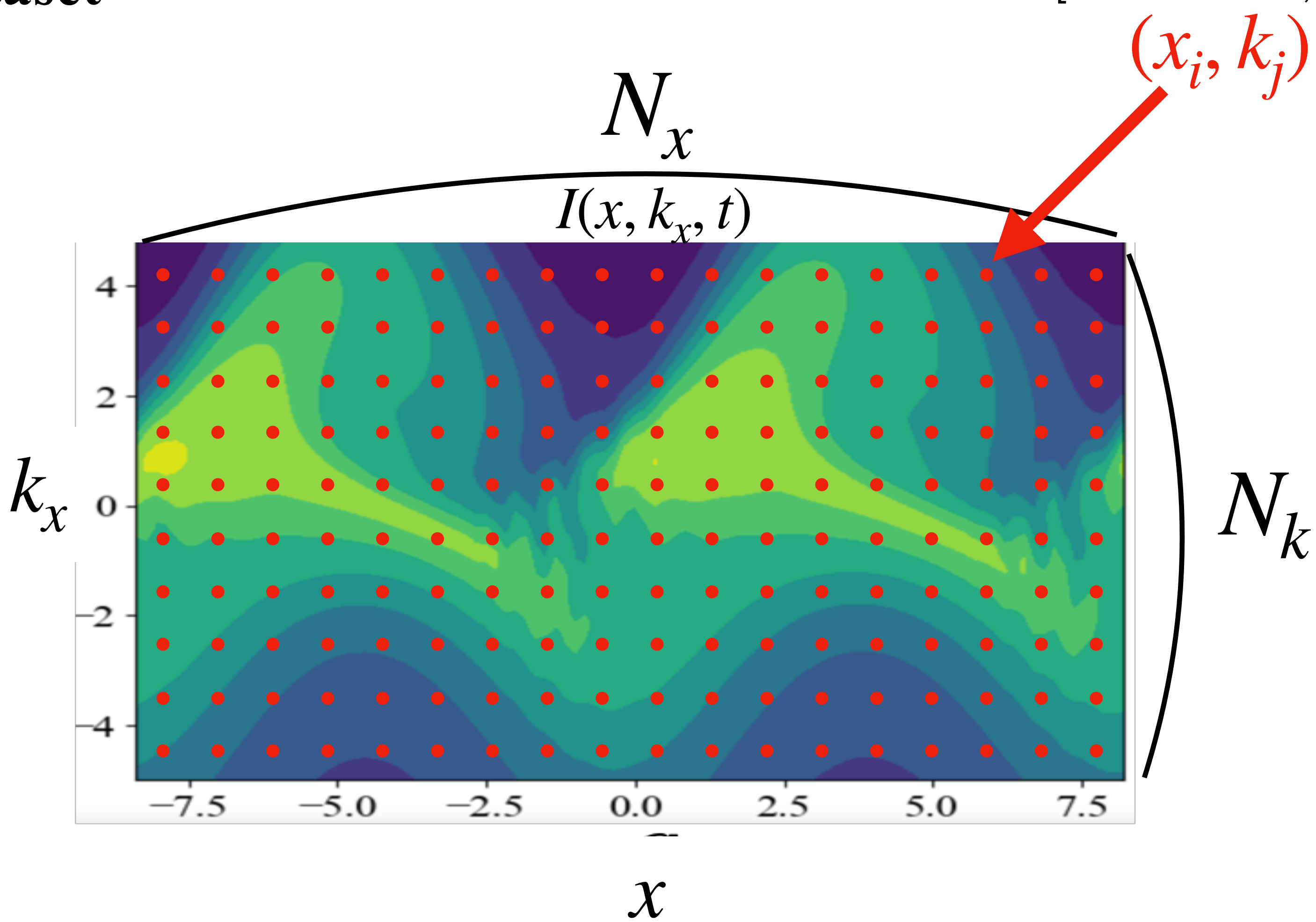
[Y. Mototake, M. Sasaki, arXiv:2511.04564]



Proposed framework

○ **Setup of training dataset**

[Y. Mototake, M. Sasaki, arXiv:2511.04564]



$\partial_t I(x, k_x, t), \partial_x I(x, k_x, t), \partial_{k_x} I(x, k_X, t), C(x, k_x, t)$ is **calculated numerically** from $I(x, k, t)$

↑We know the true Hamiltonian

【Dataset】

$k := k_x$

$$D = \left\{ \partial_t I(x_i, k_j, t), \partial_x I(x_i, k_j, t), \partial_k I(x_i, k_j, t), C(x_i, k_j, t) \,\middle|\, i \in [0, N_x], j \in [0, N_k], t = t_1 \right\}$$

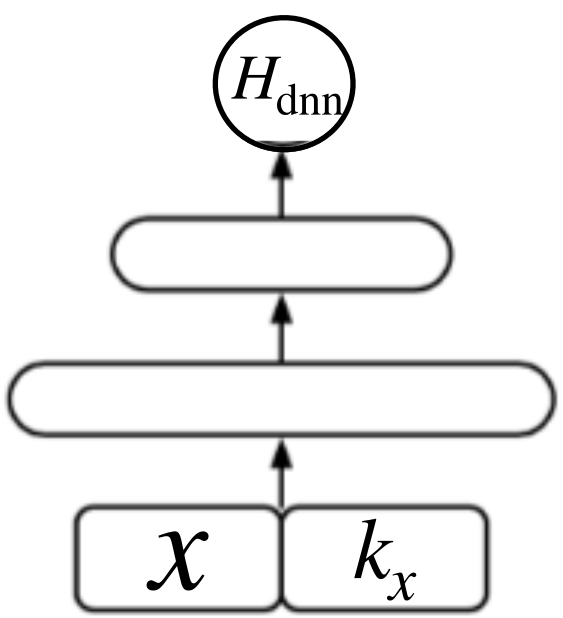
Proposed framework

○ **Loss function of DNN Hamiltonian**

[Y. Mototake, M. Sasaki, arXiv:2511.04564]

【Dataset】

$$D = \left\{ \partial_t I(x_i, k_j, t), \partial_x I(x_i, k_j, t), \partial_k I(x_i, k_j, t), C(x_i, k_j, t) \,\middle|\, i \in [0, N_x], j \in [0, N_k], t = \tau \right\}$$



Time derivation of $I(x, k, t)$ following the DNN Hamiltonian H_{dnn}

$$\partial_t I(x, k, t) = -\partial_k H_{\mathrm{dnn}}(x, k, t)\partial_x I(x, k, t) + \partial_x H_{\mathrm{dnn}}(x, k, t)\partial_k I(x, k, t) + C(x, k, t) \Big|_{k := k_x}$$

【Loss function of DNN】

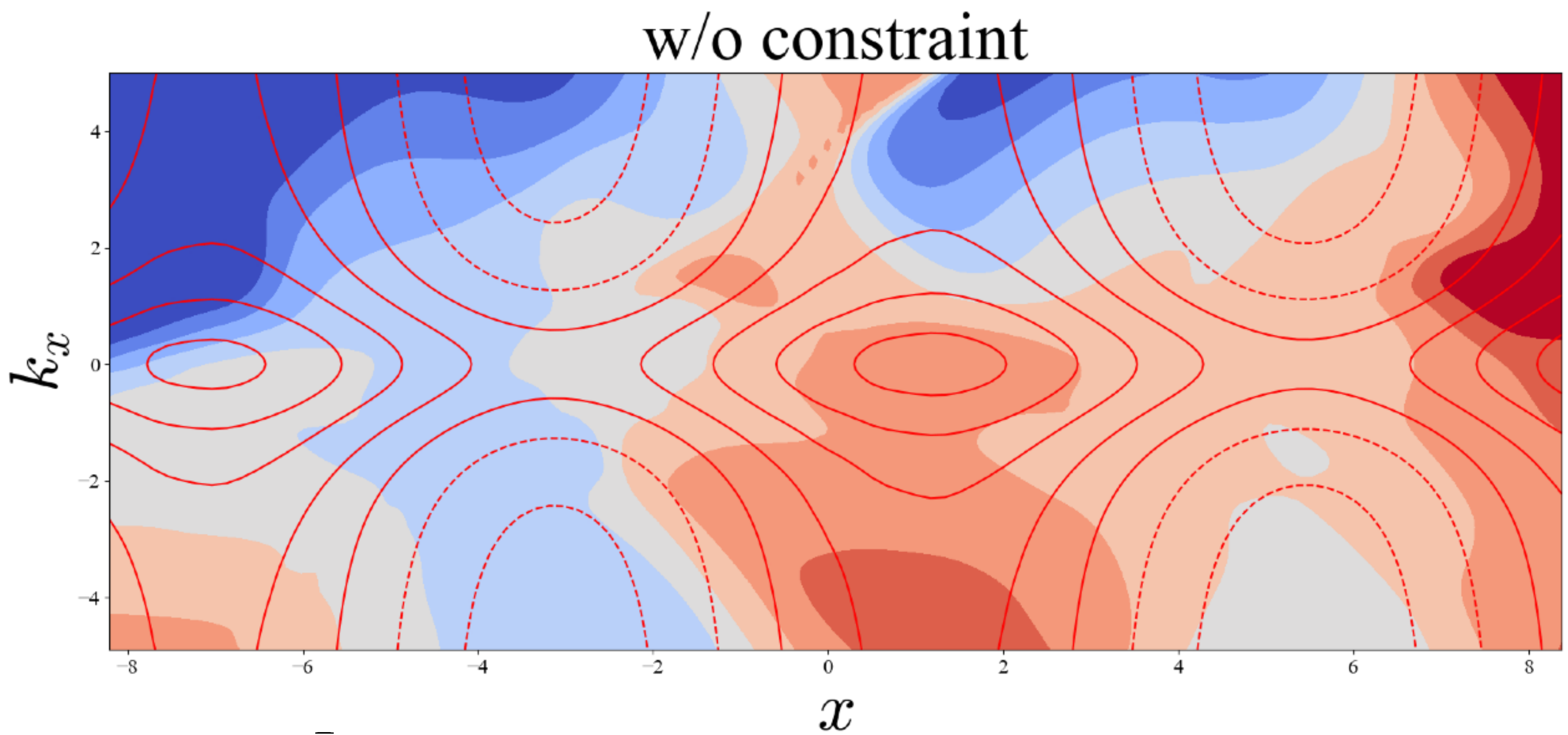
(Learn Hamiltonians at certain time t)

$$\mathrm{Loss}(\mathbf{w}_{\mathrm{dnn}}) = \frac{1}{N_x N_{k_x}} \sum_{i,j} \left\{ \partial_t I(x_i, k_j, \tau) - \left[-\partial_x H_{\mathrm{dnn}}(x_i, k_j, \tau; \mathbf{w}_{\mathrm{dnn}}) \partial_x I(x_i, k_j, \tau) + \partial_x H_{\mathrm{dnn}}(x_i, k_j, \tau; \mathbf{w}_{\mathrm{dnn}}) \partial_k I(x_i, k_j, \tau) + C(x_i, k_j, \tau) \right] \right\}^2$$

Result

○ **Verification results for simulation dataset in which the true Hamiltonian is known**

[Y. Mototake, M. Sasaki, arXiv:2511.04564]



Red contour: True Hamiltonian
Histogram: Hamiltonian estimated by the neural network

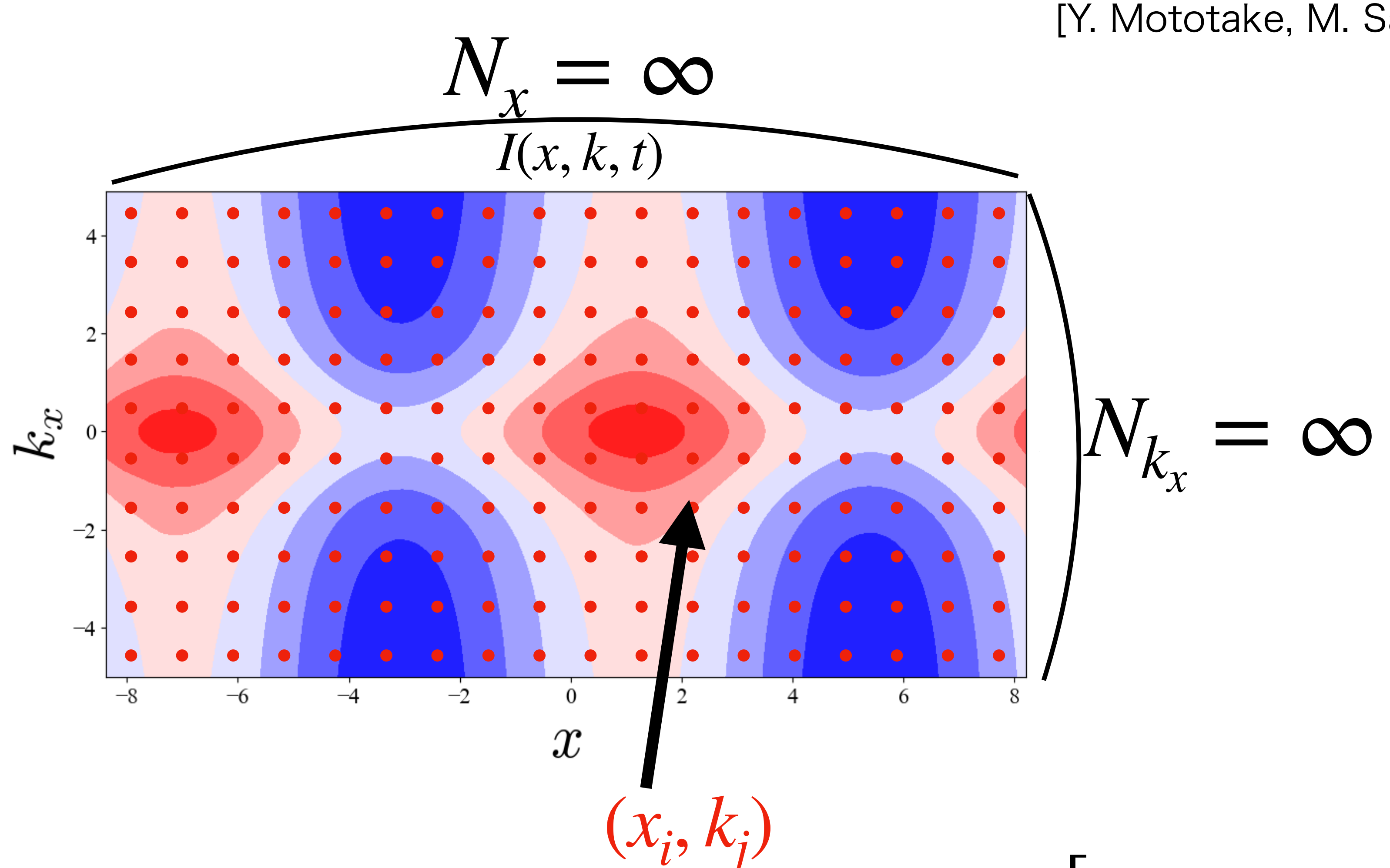
⇒**Estimated Hamiltonian has completely different property of true Hamiltonian H_{true}**

$$\frac{\partial I(x,k_x,t)}{\partial t} + \frac{\partial H(x,k_x,t)}{\partial k_x}\frac{\partial I(x,k_x,t)}{\partial x} - \frac{\partial H(x,k_x,t)}{\partial x}\frac{\partial I(x,k_x,t)}{\partial k_x} = C(x,k_x,t)$$

⇒**There are structural uncertainties** $H(x,k_x,t) = H_{\mathrm{true}}(x,k_x,t) + kf[I(x,k_x,t)]$, $k \in \mathbb{R}$, f : **Arbitrary differentiable function.**

○ **Evaluation of structural uncertainties**

[Y. Mototake, M. Sasaki, arXiv:2511.04564]



$k := k_x$

$$H(x_i, k_j, \tau) - H(x_0, k_0, \tau) = \left[H(x_i, k_0, \tau) - H(x_0, k_0, \tau)\right] + \left[H(x_i, k_j, \tau) - H(x_i, k_0, \tau)\right]$$

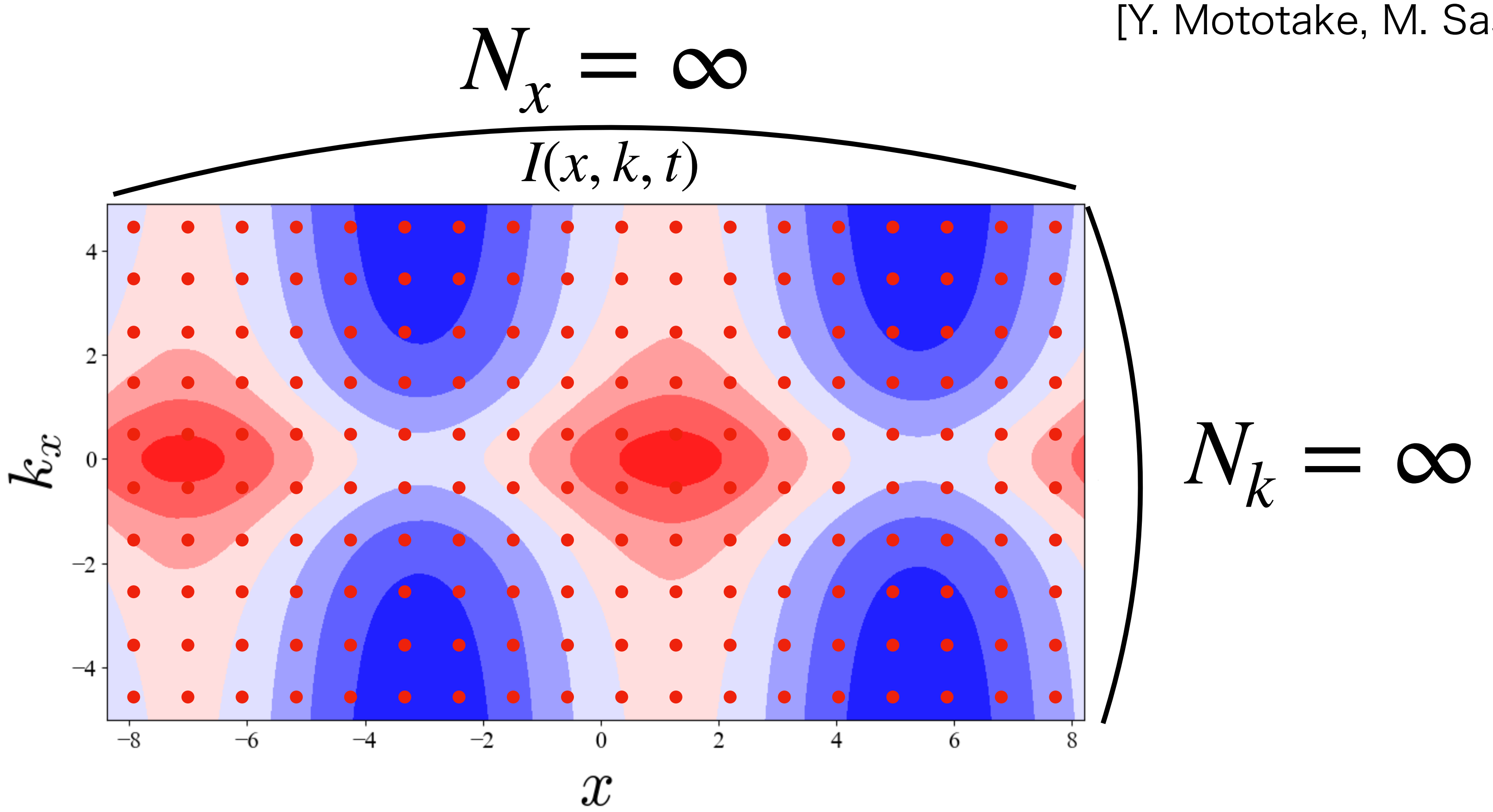
$$= \int_{x_0}^{x_i} dx \partial_x H(x, k_0, \tau) + \int_{k_0}^{k_j} dk \partial_k H(k, x, \tau)$$

$$= x_0 + \lim_{\Delta_x \to 0} \sum_{k=1}^{i} \partial_x H(x_0 + k\Delta_x, k_0, \tau)\Delta_x + k_0 + \lim_{\Delta_k \to 0} \sum_{l=1}^{j} \partial_k H(x, k_0 + l\Delta_k, \tau)\Delta_k$$

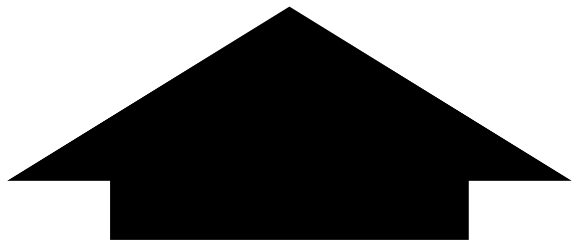
▶**If all partial differential coefficients, $\partial_x H(x_i, k_j)$, $\partial_k H(x_i, k_j)$, on the infinitesimally small interval grid are determined, the Hamiltonian $H(x, k, \tau)$ is determined except for the constants**

○ **Evaluation of structural uncertainties**

[Y. Mototake, M. Sasaki, arXiv:2511.04564]



$$\partial_t I(x_i, k_j, \tau) = -\partial_k H_{\mathrm{dnn}}(x_i, k_j, \tau)\partial_x I(x_i, k_j, \tau) + \partial_x H_{\mathrm{dnn}}(x_i, k_j, \tau)\partial_k I(x_i, k_j, \tau) + C(x_i, k_j, \tau)$$

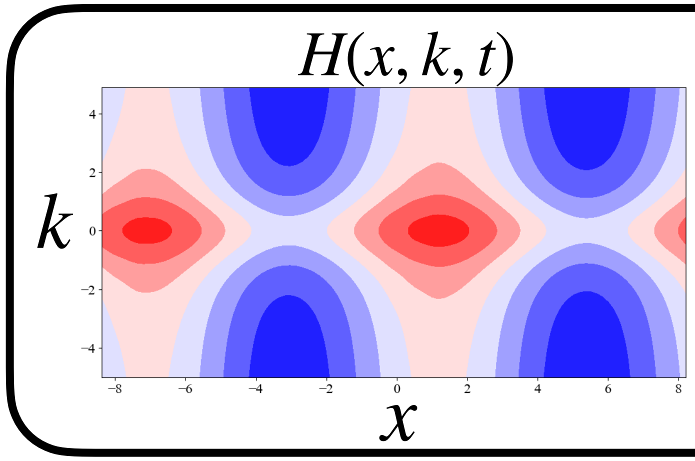


<u>The source of uncertainties lies in attempting to estimate two micro-coefficients using a single equation.</u>

○ Relaxing uncertainties by Introducing physical constraints

[Y. Mototake, M. Sasaki, arXiv:2511.04564]

$$
\begin{pmatrix} I_t(x_1,k_1) \\ \vdots \\ I_t(x_1,k_{N_k}) \\ I_t(x_2,k_1) \\ \vdots \\ I_t(x_{N_x},k_{N_k}) \end{pmatrix}
=
\underbrace{\begin{pmatrix} -I_k(x_1,k_1) & I_x(x_1,k_1) & 0 & 0 & & \\ 0 & 0 & -I_k(x_1,k_2) & I_x(x_1,k_2) & & \mathbf{O} \\ & \mathbf{O} & & & \ddots & \\ & & & & -I_k(x_{N_x},k_{N_k}) & I_x(x_{N_x},k_{N_k}) \end{pmatrix}}_{N_qN_p \times 2N_qN_p \ \textbf{matrix}}
\begin{pmatrix} H_x(x_1,k_1) \\ H_k(x_1,k_1) \\ \vdots \\ H_x(x_{N_x},k_{N_k}) \\ H_k(x_{N_x},k_{N_k}) \end{pmatrix}
$$



←symmetric

$$H\left(x^{(i)}, -k_x^{(j)}, t^{(k)}\right) = H\left(x^{(i)}, k_x^{(j)}, t^{(k)}\right)$$

$$\partial_x H\left(x^{(i)}, -k_x^{(j)}, t^{(k)}\right) = \partial_x H\left(x^{(i)}, k_x^{(j)}, t^{(k)}\right)$$

$$\partial_k H\left(x^{(i)}, -k_x^{(j)}, t^{(k)}\right) = -\partial_k H\left(x^{(i)}, k_x^{(j)}, t^{(k)}\right)$$

$$
\begin{pmatrix} I_t(x_1,k_1) \\ I_t(x_1,k_2) \\ \vdots \\ I_t(x_2,k_{N_{k_x}}) \\ I_t(x_{N_x},k_{N_{k_x}}) \end{pmatrix}
=
\underbrace{\begin{pmatrix} -I_k(x_1,k_1) & I_x(x_1,k_1) & & & \\ -I_k(x_1,-k_1) & -I_x(x_1,-k_1) & & \mathbf{O} & \\ & & \ddots & & \\ & & & \ddots & \\ & \mathbf{0} & & -I_k(x_{N_x/2},k_{N_k/2}) & I_x(x_{N_x/2},k_{N_k/2}) \\ & & & -I_p(x_{N_x/2},-k_{N_k/2}) & -I_q(x_{N_x/2},-k_{N_k/2}) \end{pmatrix}}_{N_qN_p \times N_qN_p \ \textbf{matrix}}
\begin{pmatrix} H_x(x_1,k_1) \\ H_k(x_1,k_1) \\ \vdots \\ H_x(x_{N_x/2},k_{N_k/2}) \\ H_k(x_{N_x/2},k_{N_k/2}) \end{pmatrix}
$$

▶The number of equations coincides with the number of partial differential coefficients

○ **Relaxing uncertainties by Introducing physical constraints**

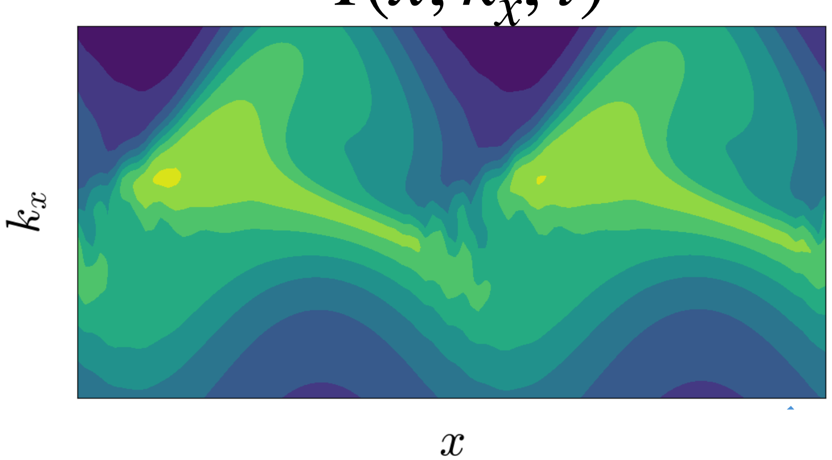
[Y. Mototake, M. Sasaki, arXiv:2511.04564]

$$
\begin{pmatrix} I_t(x_1,k_1) \\ I_t(x_1,k_2) \\ \vdots \\ I_t(x_2,k_{N_{k_x}}) \\ I_t(x_{N_x},k_{N_{k_x}}) \end{pmatrix}
=
\underbrace{\begin{pmatrix} -I_k(x_1,k_1) & I_x(x_1,k_1) & & & \\ -I_k(x_1,-k_1) & -I_x(x_1,-k_1) & & \mathbf{O} & \\ & & \ddots & & \\ & \mathbf{O} & & -I_k(x_{N_x/2},k_{N_k/2}) & I_x(x_{N_x/2},k_{N_k/2}) \\ & & & -I_p(x_{N_x/2},-k_{N_k/2}) & -I_q(x_{N_x/2},-k_{N_k/2}) \end{pmatrix}}_{A}
\begin{pmatrix} H_x(x_1,k_1) \\ H_k(x_1,k_1) \\ \vdots \\ H_x(x_{N_x/2},k_{N_k/2}) \\ H_k(x_{N_x/2},k_{N_k/2}) \end{pmatrix}
$$

If A is full rank $\leftrightarrow \forall\, i,j$, rank $\left[\begin{pmatrix} -I_k(x_i,k_j) & I_x(x_i,k_j) \\ -I_k(x_i,-k_j) & -I_x(x_i,-k_j) \end{pmatrix}\right] = 2,$

then Hamiltonian is determined: $\begin{pmatrix} H_x(x_1,k_1) \\ H_k(x_1,k_1) \\ \vdots \\ H_x(x_{N_x/2},k_{N_k/2}) \\ H_k(x_{N_x/2},k_{N_k/2}) \end{pmatrix} = A^{-1} \begin{pmatrix} I_t(x_1,k_1) \\ I_t(x_1,k_2) \\ \vdots \\ I_t(x_2,k_{N_k}) \\ I_t(x_{N_x},k_{N_k}) \end{pmatrix}.$

$I(x, k_x, t)$



$$(-I_k(x_i,k_j), I_x(x_i,k_j)) \neq s(-I_k(x_i,-k_j), I_x(x_i,-k_j)),$$

$s \in \mathbb{R}.$

⬇

A is a full rank!!

▶ Structural uncertainties are disappeared

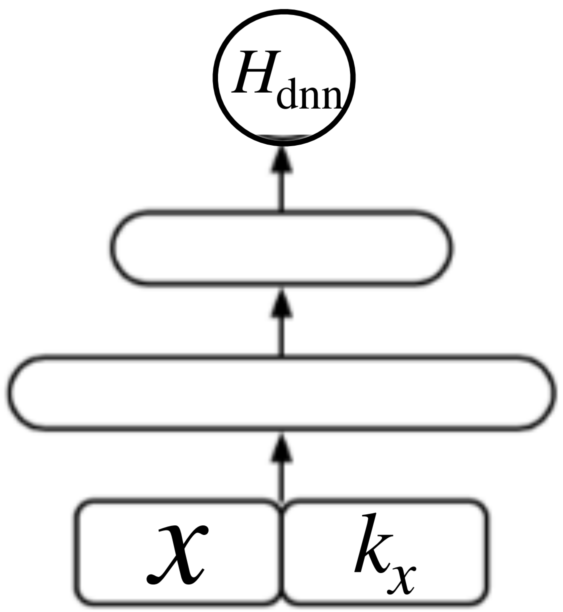
Proposed framework (improved)

○ **Loss function of DNN Hamiltonian**

[Y. Mototake, M. Sasaki, arXiv:2511.04564]

【Dataset】

$$D = \left\{ \partial_t I(x_i, k_j, t), \partial_x I(x_i, k_j, t), \partial_k I(x_i, k_j, t), C(x_i, k_j, t) \,\middle|\, i \in [0, N_x], j \in [0, N_k], t = \tau \right\}$$



Time derivation of $I(x, k, t)$ following the DNN Hamiltonian H_{dnn}

$$\partial_t I(x,k,t) = -\partial_k H_{\mathrm{dnn}}(x,k,t)\partial_x I(x,k,t) + \partial_x H_{\mathrm{dnn}}(x,k,t)\partial_k I(x,k,t) + C(x,k,t)$$

【Loss function of DNN】

$$\mathrm{Loss}(\mathbf{w}_{\mathrm{dnn}}) = \frac{1}{N_x N_{k_x}} \sum_{i,j} \left\{ \partial_t I(x_i, k_j, \tau) - \left[-\partial_x H_{\mathrm{dnn}}(x_i, k_j, \tau; \mathbf{w}_{\mathrm{dnn}}) \partial_x I(x_i, k_j, \tau) + \partial_x H_{\mathrm{dnn}}(x_i, k_j, \tau; \mathbf{w}_{\mathrm{dnn}}) \partial_k I(x_i, k_j, \tau) + C(x_i, k_j, \tau) \right] \right\}^2$$

Constraint of symmetry→

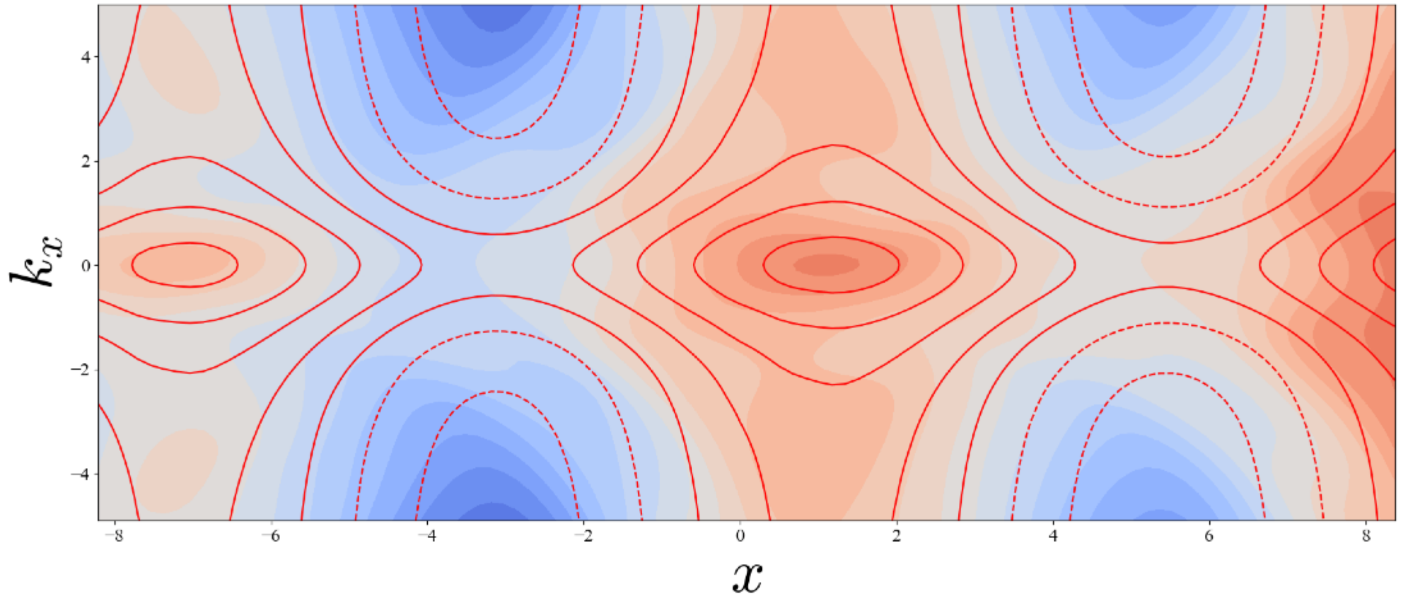
$$+\lambda \sum_{i,j} \left[\left| H_{\mathrm{dnn}}(x_i, k_j, t_k; \mathbf{w}_{\mathrm{dnn}}) - H_{\mathrm{dnn}}(x_i, -k_j, t_k; \mathbf{w}_{\mathrm{dnn}}) \right|^2 \right]$$

Results (improved)

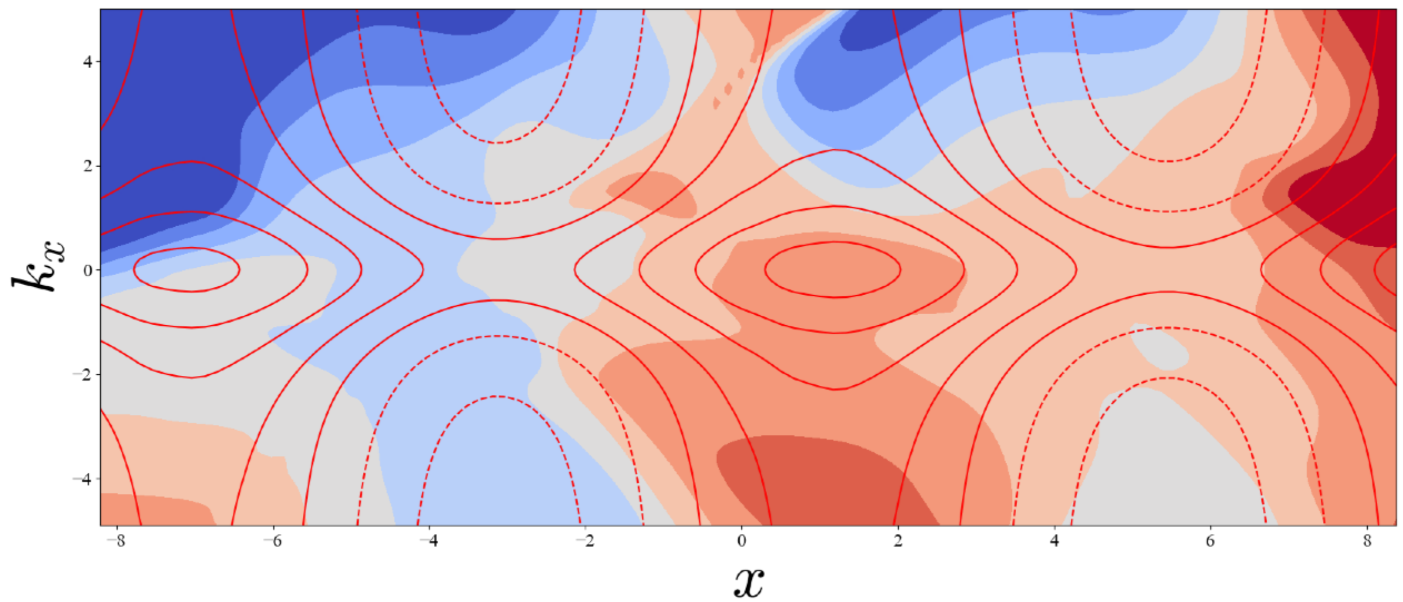
○ **Verification results for simulation dataset in which the true Hamiltonian is known**

[Y. Mototake, M. Sasaki, arXiv:2511.04564]

w constraint



w/o constraint



Red contour: True Hamiltonian
Histogram: Hamiltonian estimated by the neural network

⇒Hamiltonians are successfully estimated！！

Remark

- **1-d diffusion system**

The one-dimensional diffusion equation
$\partial_t u(t,x) = \partial_x\big(a(x)\,\partial_x u(t,x)\big)$,
where $a(x)$ is an unknown diffusion coefficient.
Expanding the right-hand side of diffusion equation gives $\partial_t u(t,x) = \partial_x a(x)\,\partial_x u(t,x) + a(x)\,\partial_{xx} u(t,x)$.
From the viewpoint of the coefficient function a, this PDE contains the first derivative $\partial_x a(x)$ (order $k-1=1$) with coefficient $\phi(x) = \partial_x u(t,x)$, in addition to the zeroth-order term $a(x)$ with coefficient $\partial_{xx} u(t,x)$.
We discretize the spatial domain on an infinitesimal grid $x_j = x_0 + jh$ ($j = 1,\ldots,N$) with spacing $h > 0$.
On this grid, we denote $s_j := \partial_x a(x_j), \qquad t_j := a(x_j)$.
Along the one-dimensional grid, the first derivative can be represented as a discrete integral (cumulative sum) of the second derivative: $t_j \;=\; t_0 \;+\; h\sum_{i=1}^{j} s_i, \qquad j = 1,\ldots,N,$ where t_0 is an integration constant corresponding to a lower-degree component of $a(x)$.
Then, the equation on an infinitesimal N grid space is written as $\mathbf{M}\cdot\mathbf{s} = \mathbf{c}$, where $\mathbf{s} = (s_1,\ldots,s_N)^\top$,
$\mathbf{c} = \big(\partial_t u(t,x_1) - t_0\partial_{xx}u(t,x_1), \ldots, \partial_t u(t,x_N) - t_0\partial_{xx}u(t,x_N)\big)$, and

$$M_{ji} = \begin{cases} h\,\partial_{xx} u(t,x_j), & i<j, \\ [2pt] h\,\partial_{xx} u(t,x_j) + \partial_x u(t,x_j), & i=j, \\ [2pt] 0, & i>j\,. \end{cases}$$

If each element of M is nonzero, M is a lower triangular matrix and $\mathrm{rank}(M) = N$.
Thus, the diffusion coefficient function $a(x)$ is uniquely determined from the diffusion equation, except for constant term.

Summary and Discussion

- The procedure used to evaluate uncertainty in physical model estimation in this study could be developed into a framework for evaluating uncertainty in a wider range of physical models.
- It is possible to estimate the Hamiltonian function of the wave kinetic equation in a data-driven manner by introducing appropriate constraints.
- Potential to develop a framework for estimating the Hamiltonian function from experimental data.
- The proposed framework may provide a framework for investigating the correspondence between wave kinetic theory and other turbulence models that reproduce zonal flows.

[Y. Mototake, M. Sasaki, arXiv:2511.04564]

Importance of Uncertainty Modeling

Maxwell equation

$$\mathbf{B} = \nabla \times \mathbf{A}$$

$$\mathbf{A} \to \mathbf{A} + \nabla f$$

⇒ Gauge theory

⇒ The structure of the indeterminate equivalence class itself is important interpretable information

De Rham cohomology

Definition [edit]

The **de Rham complex** is the cochain complex of differential forms on some smooth manifold M, with the exterior derivative as the differential:

$$0 \to \Omega^0(M) \stackrel{d}{\to} \Omega^1(M) \stackrel{d}{\to} \Omega^2(M) \stackrel{d}{\to} \Omega^3(M) \to \cdots,$$

where $\Omega^0(M)$ is the space of smooth functions on M, $\Omega^1(M)$ is the space of 1-forms, and so forth. Forms that are the image of other forms under the exterior derivative, plus the constant 0 function in $\Omega^0(M)$, are called **exact** and forms whose exterior derivative is 0 are called **closed** (see *Closed and exact differential forms*); the relationship $d^2 = 0$ then says that exact forms are closed.

$$H^k_{\mathrm{dR}}(M) = \mathrm{Ker}\, d_k / \mathrm{Im}\, d_{k-1}$$

⇒Homology is important for interpretable AI in physics

⇒Topological view is important for data analysis of physics data using interpretable AI