



# 解釈可能AIによる データ駆動理学の実現へ向けた取り組み

研究室HP:

<https://mototakelab.github.io/>

最新版スライド:

<https://mototakelab.github.io/slides/measurement-informatics-seminar.pdf>

一橋大学大学院 ソーシャル・データサイエンス研究科  
准教授 本武 陽一



# 1. 自己紹介

## 2. 解釈可能AIとデータ駆動理学

## 3. DNN型アプローチによるデータ駆動理学の実現

## 4. 基底関数型アプローチによるデータ駆動理学の実現

## 5. まとめと今後の展望

# 自己紹介

福岡県福岡市出身

## 学歴：

2008/3 東北大学 理学部 物理学科卒業

2010/3 北海道大学 大学院理学院 宇宙理学専攻

修士課程修了 指導教官：石川健三 教授（素粒子物理学、テーマ：波束のダイナミクス）

2013/3 東京大学 大学院総合文化研究科 広域科学専攻

修士課程修了 指導教官：植田一博 教授（認知科学、テーマ：脳活動を用いたHAI評価）

2016/3 東京大学 大学院総合文化研究科 広域科学専攻

博士課程修了 指導教官：池上高志 教授（複雑系、テーマ：深層学習、群れ）

## 職歴：

2016/5 ~ 2019/3 東京大学 大学院新領域創成科学研究科 複雑理工学専攻

特任研究員（岡田真人教授、テーマ：データ駆動科学）

2019/4 ~ 統計数理研究所 統計的機械学習研究センター

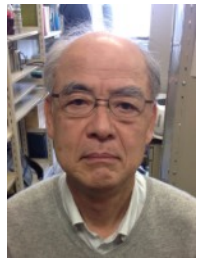
特任助教（福水健次教授、テーマ：統計的機械学習）

2023/1 ~ 一橋大学大学院ソーシャルデータサイエンス研究科

准教授（テーマ：データ駆動理学、解釈可能AI、機械学習）

 「解釈可能AIによるパターンダイナミクスの数理構造抽出と材料情報学への応用」 研究代表

 「革新的セラミックス材料設計のための材料パターン情報学の創成」 業務管理者（研究代表）





2023年度新設

## SOCIAL DATA SCIENCE

ソーシャル・データサイエンス学部・研究科

### 学部

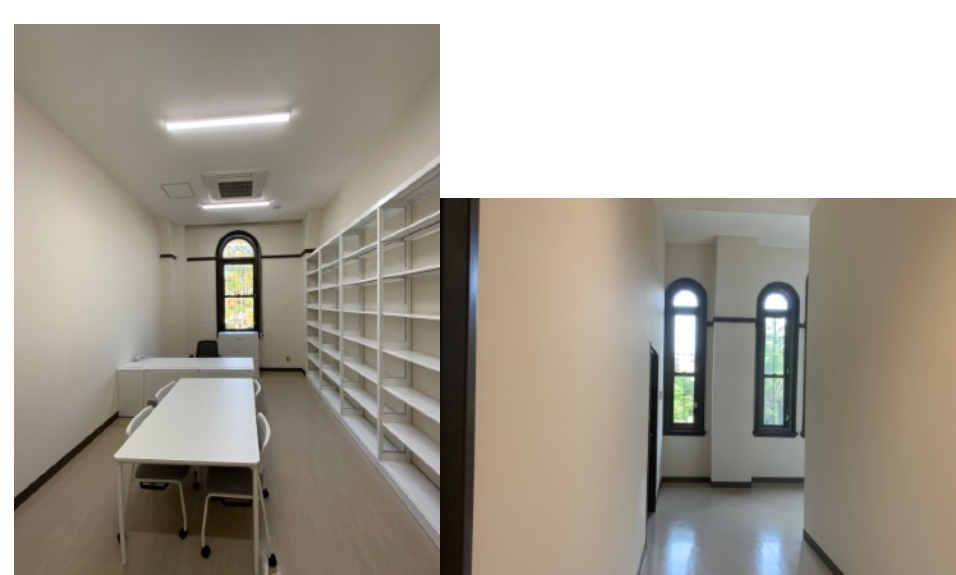
➤ 入学定員 60名(収容定員240名)

### 研究科

➤ 修士課程入学定員 21名(収容定員42名)

### 担当予定教員一覧

	教員名			研究キーワード
社会科学分野	*	七丈 直弘	教授	科学技術政策、 <b>計算材料科学</b> 、技術経営論
	*	寺田 麻佑	教授	先端技術と法、行政法、情報法
	*	檜山 敦	教授	<b>人間拡張工学、バーチャルリアリティ</b> ジェロンテクノロジー
	*	勝又 裕斗	准教授	政治学方法論、計量政治学
	*	永山 晋	准教授	経営学、組織論、創造性、ウェルビーイング
		清家 大嗣	助教	<b>情報学</b> 、ブロックチェーン、分散台帳技術、分散型 ID
統計学分野	*	今井 晋	教授	労働経済学、産業組織論、応用計量経済学
	*	清水 千弘	教授	指数理論、応用計量経済学、 <b>多変量解析</b>
	*	渡部 敏明	教授	計量ファイナンス、マクロ計量経済学、 <b>ベイズ</b> 計量経済学
	*	植松 良公	准教授	<b>統計学</b> 、高次元データ分析、時系列分析
	*	加藤 諒	准教授	マーケティング・サイエンス、 <b>ベイズ</b> 統計、欠測データ解析
	*	城田 慎一郎	准教授	<b>ベイズ</b> 統計、空間/時空間統計、計算統計
情報・AI分野	*	小町 守	教授	<b>計算言語学</b> 、自然言語処理、人工知能
	*	鈴木 真介	教授	神経経済学、計算論的神経科学、計算論的精神医学
	*	樺 惇志	准教授	情報検索、自然言語処理、対話システム
		福田 玄明	准教授	認知科学、視覚心理学、認知神経科学
	*	本武 陽一	准教授	<b>データ駆動科学</b> 、解釈可能 AI、機械学習
	*	谷田川 達也	准教授	<b>コンピュータ・グラフィクス</b> 、三次元形状処理、画像情報処理





- 研究分野

データ駆動科学、解釈可能AI、機械学習

- 研究内容



統計量や周期性で記述できない  
複雑な時空間構造を持つ現象  
(パターンダイナミクス)

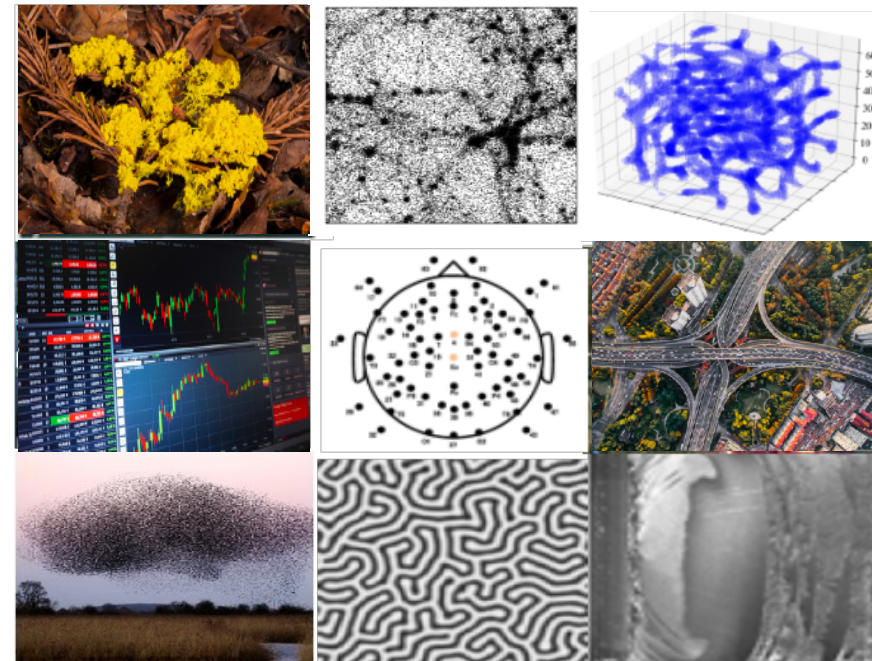
機械学習

解釈可能な情報

法則・理論

【ターゲットとする現象】

あらゆる分野のパターンダイナミクス



- 求める人物像 (以下を持っている人)

オープンマインド、現象に対する知的好奇心、気合と根性、(基本的な数学・情報科学の知識)



## 研究室スタッフ

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	天本 義史	一橋大学大学院 ソーシャル・データサイエンス研究科	高分子化学 データ駆動科学
特任助教	熊添 博之	一橋大学大学院 ソーシャル・データサイエンス研究科	物性物理 データ駆動科学
研究補助員	川田 里実	一橋大学大学院 ソーシャル・データサイエンス研究科	--
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

+ 修士課程学生 1名

1. 自己紹介

**2. 解釈可能AIとデータ駆動理学**

3. DNN型アプローチによるデータ駆動理学の実現

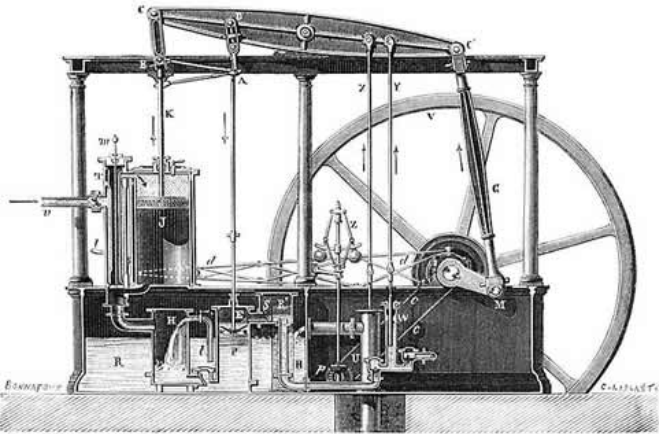
4. 基底関数型アプローチによるデータ駆動理学の実現

5. まとめと今後の展望



## 科学（理学）における抽象化と外挿性の威力

現象



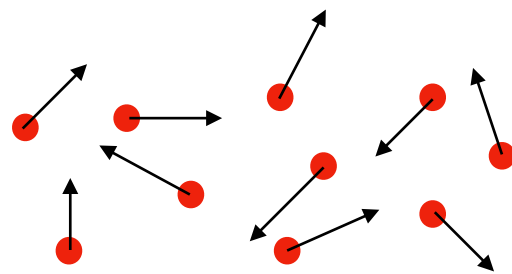
蒸気機関

実体論

現象の内挿的モデル化と予測

現象のモデル化

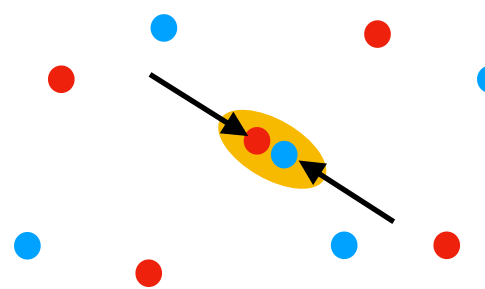
気体の分子運動論



科学的洞察

現象の予測

化学反応速度論



本質論

抽象化による大胆な外挿

一般原理の抽出

エントロピー増大則

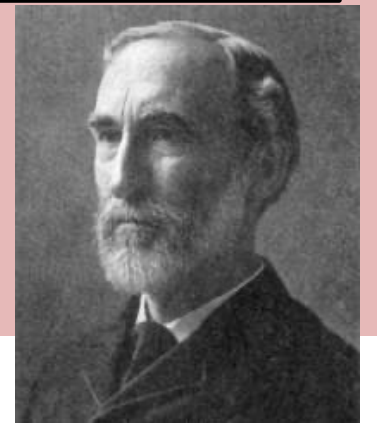
$$dS = dQ/T$$

原理の転移

化学ポテンシャル

$$\mu = \partial F / \partial N$$

Willard Gibbs



化学産業

## 解釈可能AIによる科学ループの実現=データ駆動理学

機械学習が得意

人間が得意

現象

実体論

本質論

大同特殊鋼 HPより



磁性材料

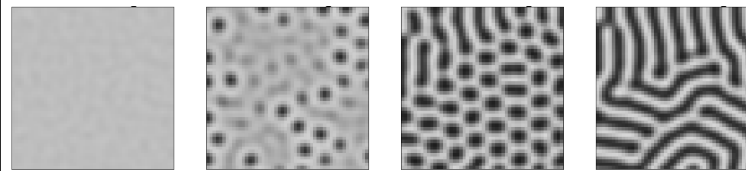
現象の内挿的モデル化と予測

抽象化による大胆な外挿

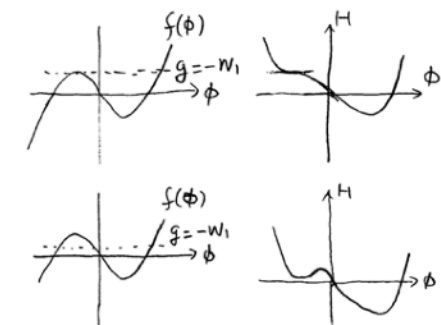
現象のモデル化

一般原理の抽出

磁区構造形成過程



パターン - 磁性特性



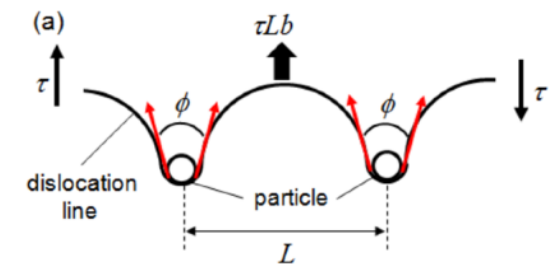
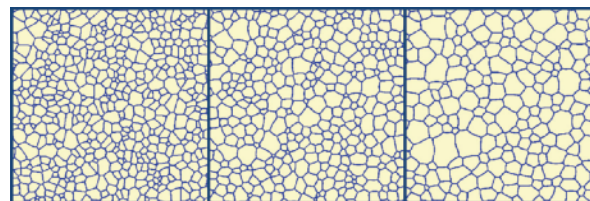
解釈可能AI

原理の転移

現象の予測

パターン - 強化機構

結晶粒形成過程



<http://www.numse.nagoya-u.ac.jp/P6/kobashi/img/file20.pdf>



構造材料



## データ駆動理学による実験計測の対応づけ

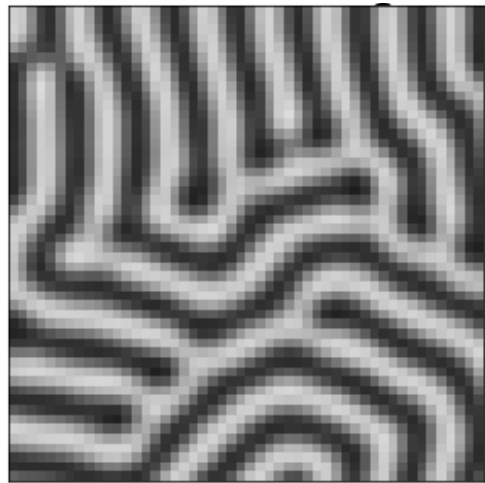
機械学習が得意

人間が得意

現象

実体論

本質論



数値計算  
現実と厳密には  
一致しない

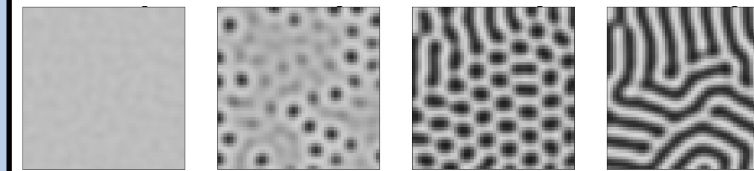
現象の内挿的モデル化と予測

抽象化による大胆な外挿

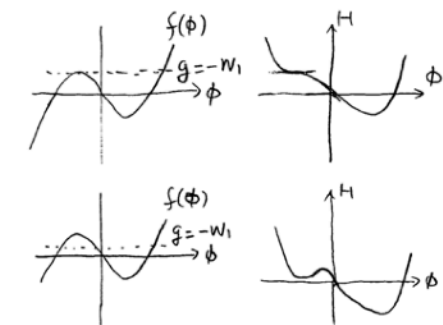
現象のモデル化

一般原理の抽出

磁区構造形成過程



パターン - 磁性特性



解釈可能AI

原理の適用

目標とする磁性特性の  
実現方法



実験・計測  
少数データ

# 解釈性をどう定義するか？

解釈性は絶対的に定義できるものか？

第236回研究会／第69回化合物新磁性材料専門研究会  
研究会 [https://www.magnetics.jp/event/topical\\_236/](https://www.magnetics.jp/event/topical_236/)

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



フーリエスペクトルの構造を”解釈できる”のは、  
計算機のない頃から計測によってスペクトルが観察できたため、  
元の空間での構造とスペクトル間の対応関係を  
経験的・体験的に会得した結果ではないか？

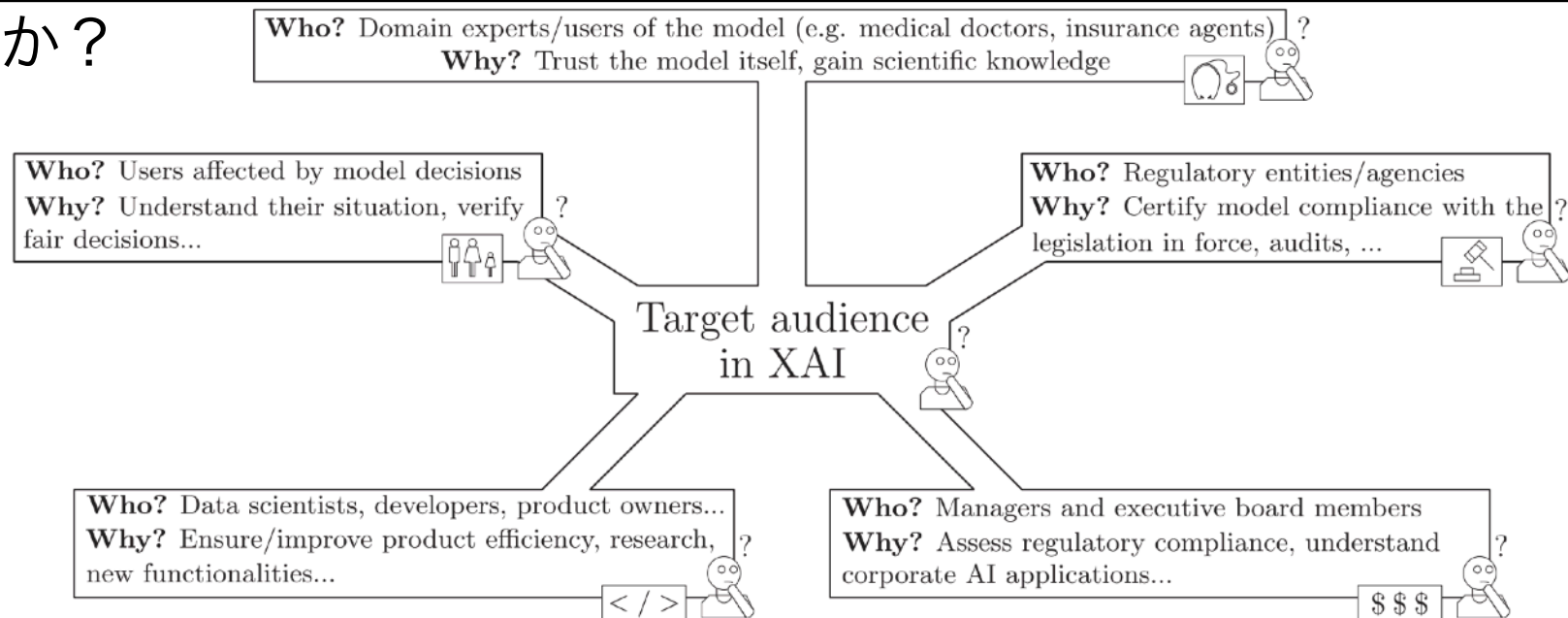


takashi Ikegami @alltbl · 5月15日

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。

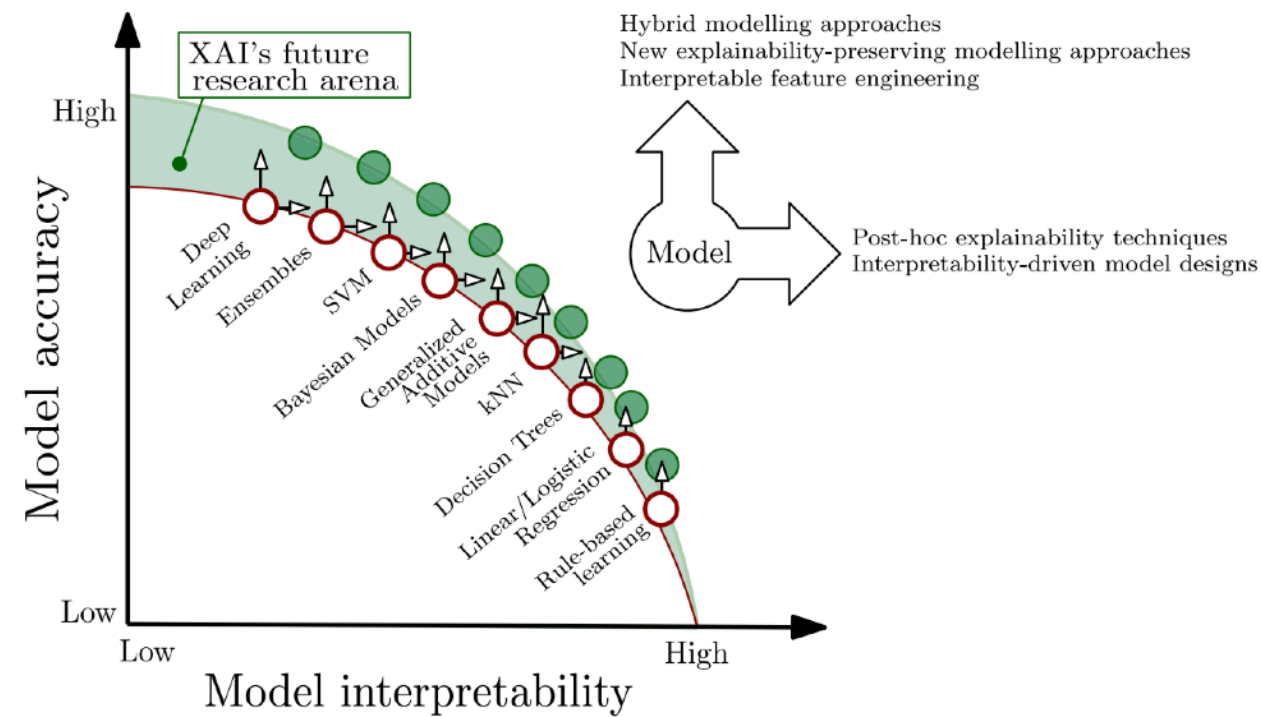
1 8 39 7,720

誰にとっての解釈か？

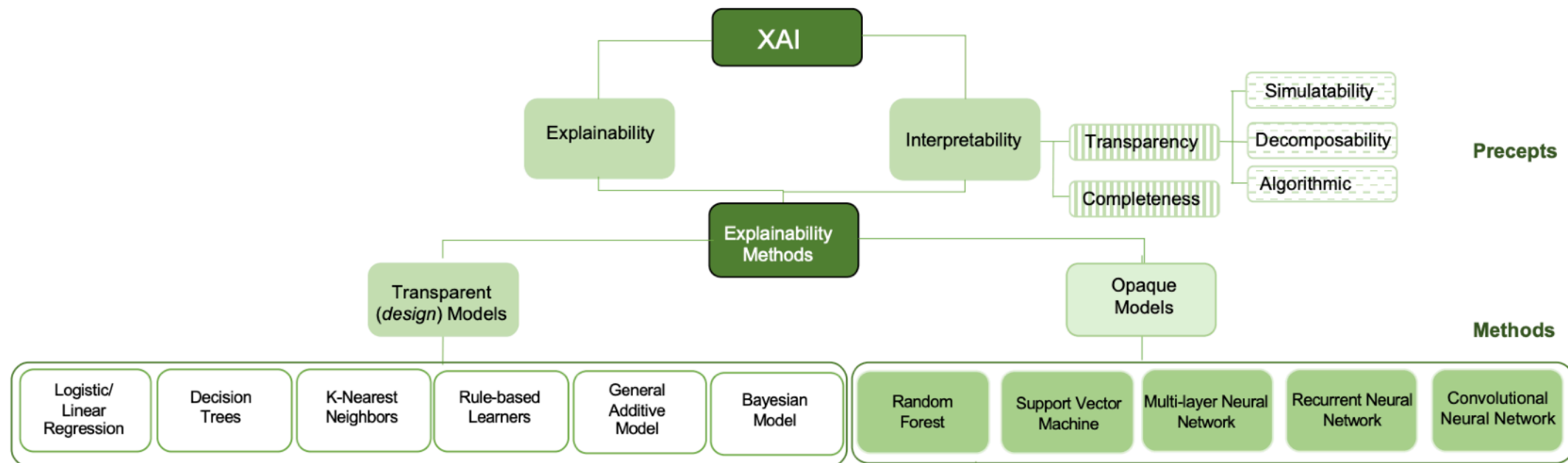


[Arrieta, Alejandro Barredo, et al., Information fusion, 58, 2020.]

▶ 「解釈性」を一般に定義することは難しい。まずは「解釈」が実現されるプロセスを観測する。



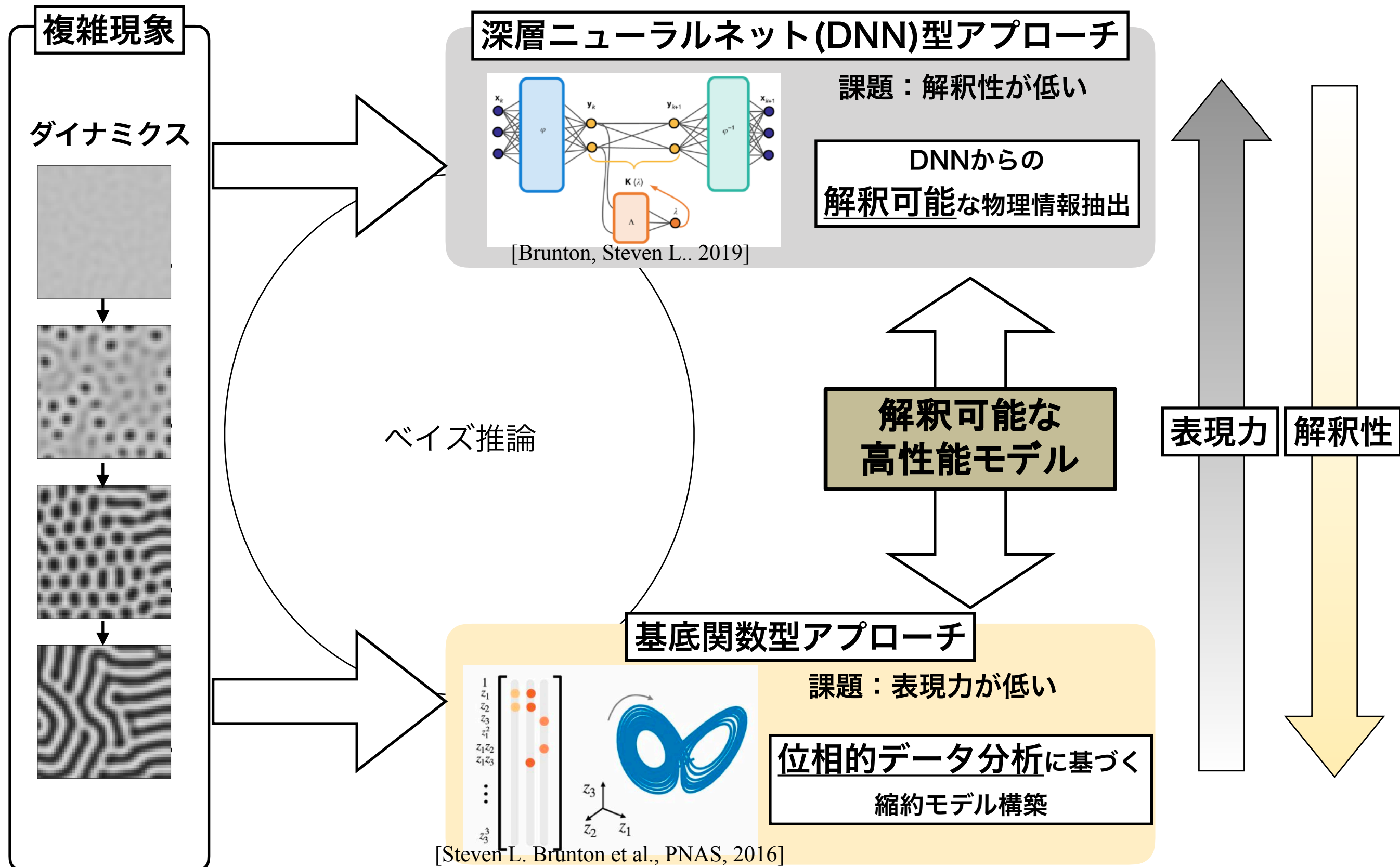
[Arrieta, Alejandro Barredo, et al., Information fusion, 58, 2020.]



[Love, Peter ED, et al., arXiv:2211.06579 2022]

▶ ブラックボックスモデルとホワイトボックスモデルで大別されたアプローチがある

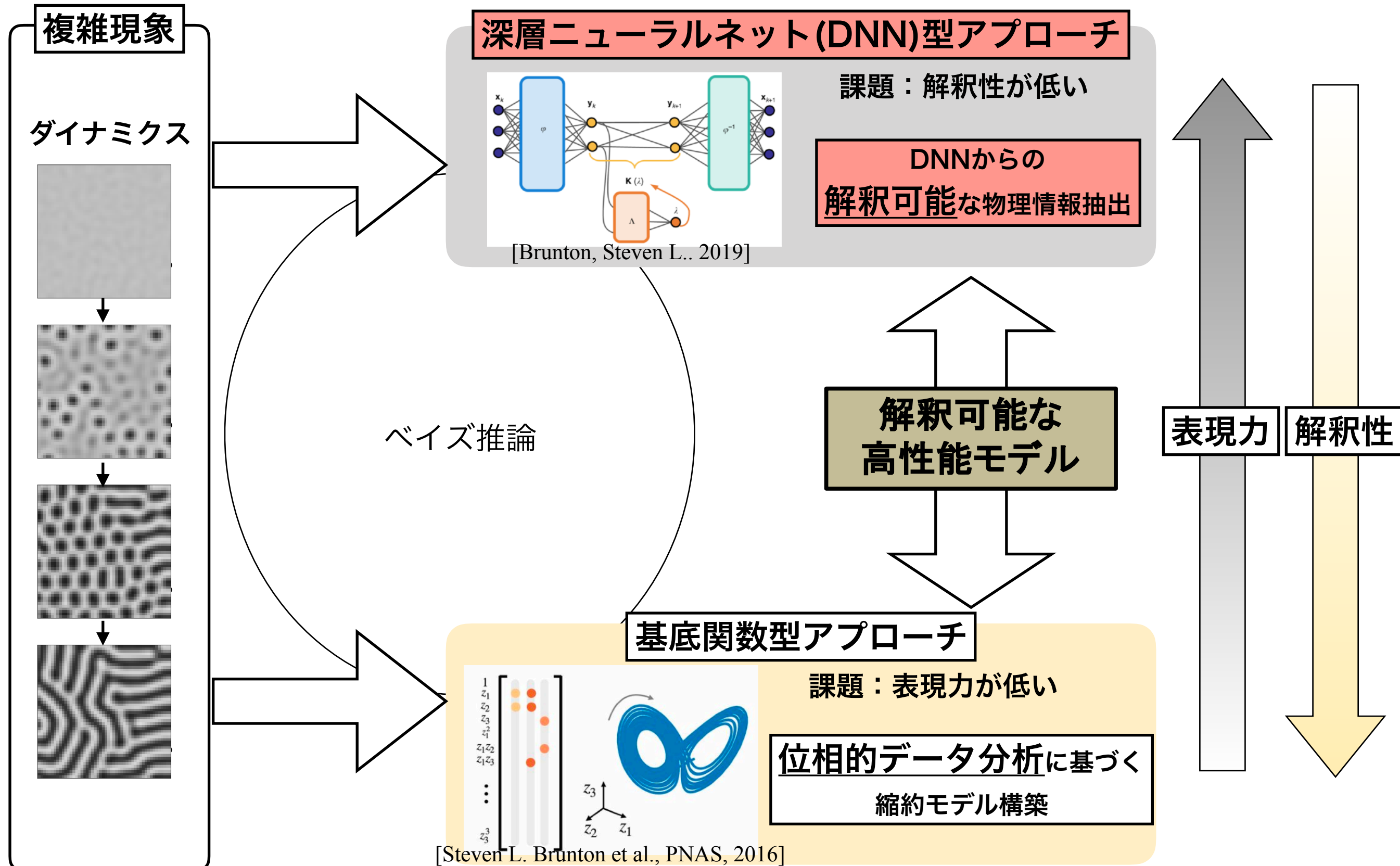
## 機械学習による解釈可能AI構築アプローチ



1. 自己紹介
2. 解釈可能AIとデータ駆動理学
- 3. DNN型アプローチによるデータ駆動理学の実現**
4. 基底関数型アプローチによるデータ駆動理学の実現
5. まとめと今後の展望

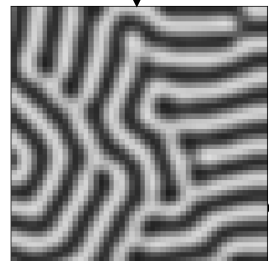
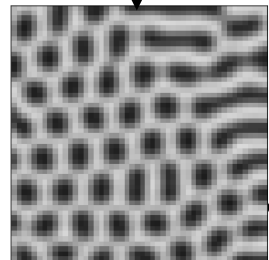
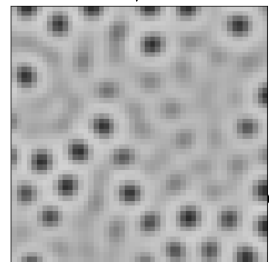
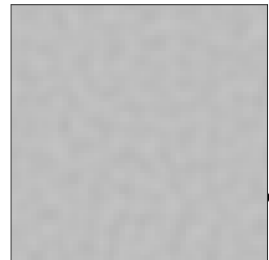


## 機械学習による解釈可能AI構築アプローチ

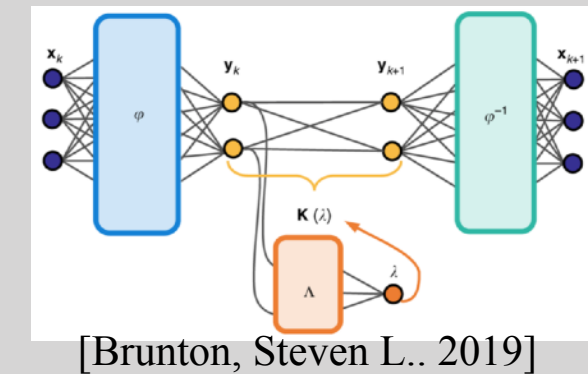


複雑現象

ダイナミクス



深層ニューラルネット (DNN) 型アプローチ



[Brunton, Steven L., 2019]

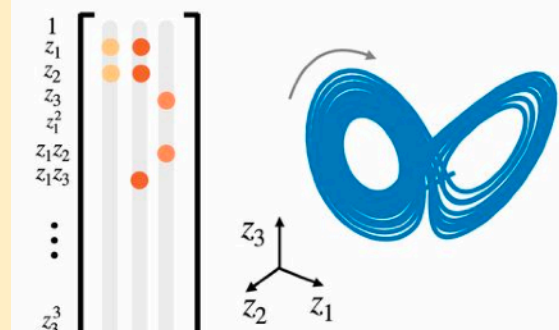
課題: 解釈性が低い

DNNからの  
解釈可能な物理情報抽出

ベイズ推論

解釈可能な  
高性能モデル

基底関数型アプローチ



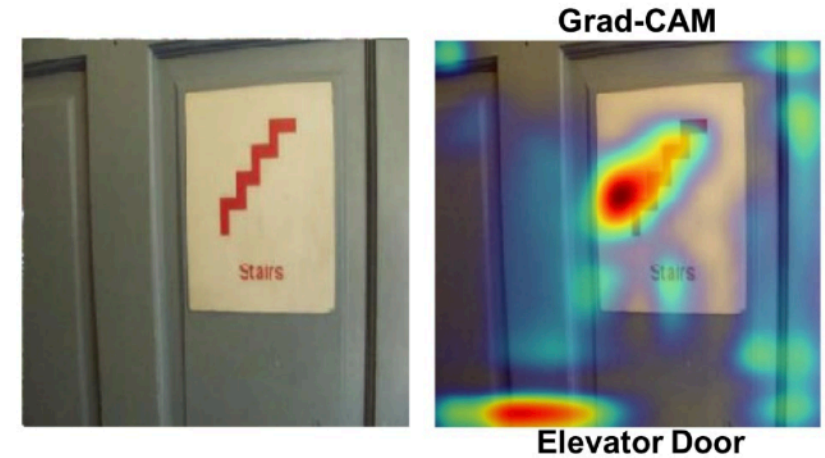
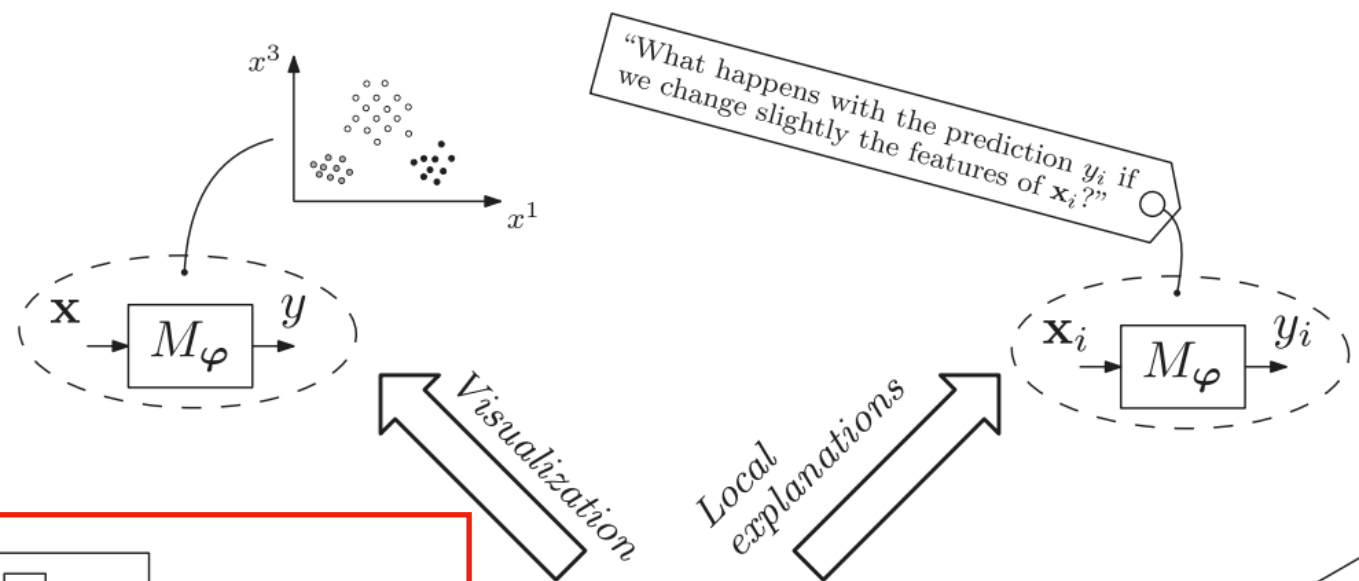
[Steven L. Brunton et al., PNAS, 2016]

課題: 表現力が低い

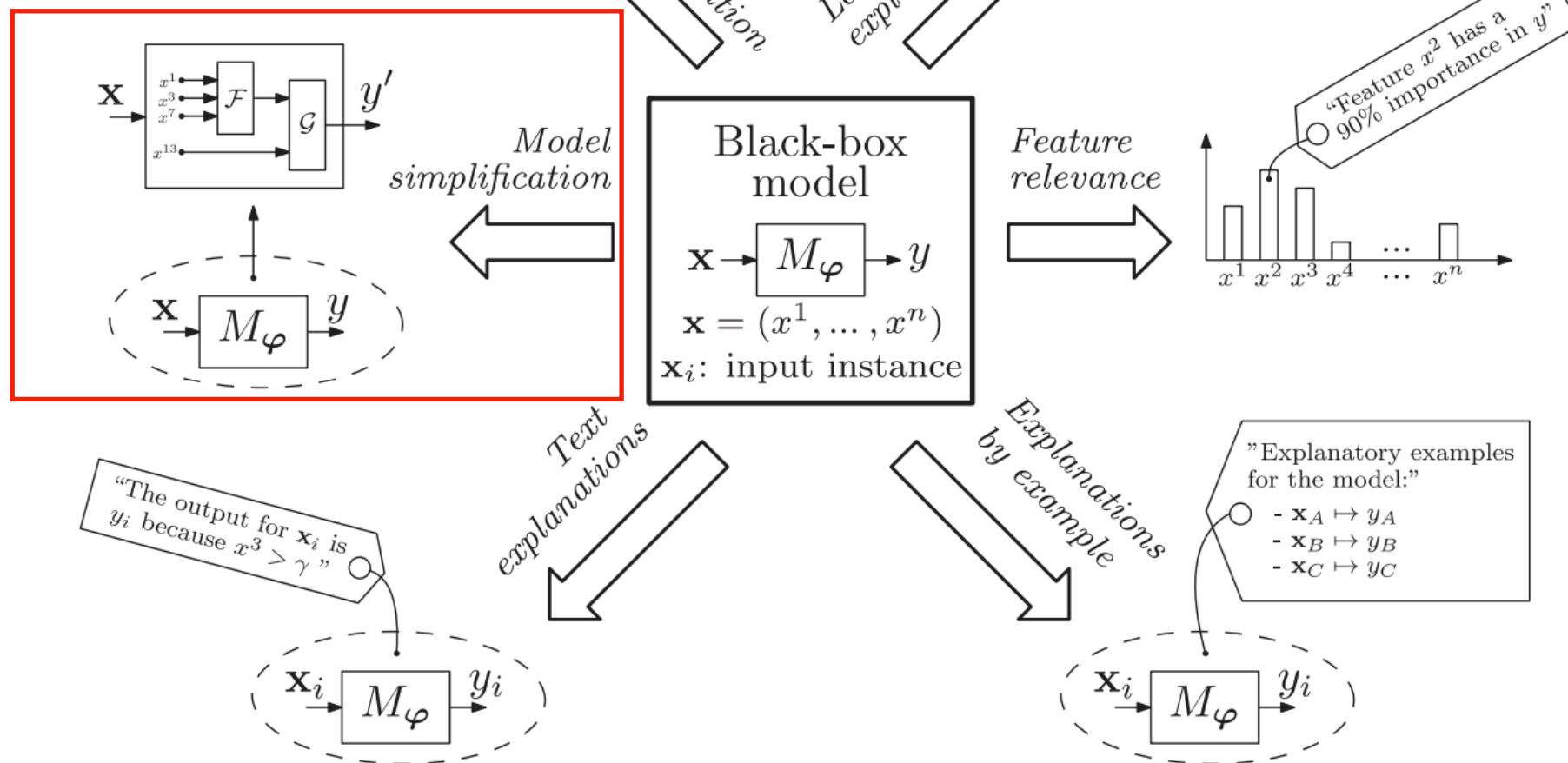
位相的データ分析に基づく  
縮約モデル構築

表現力

解釈性



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

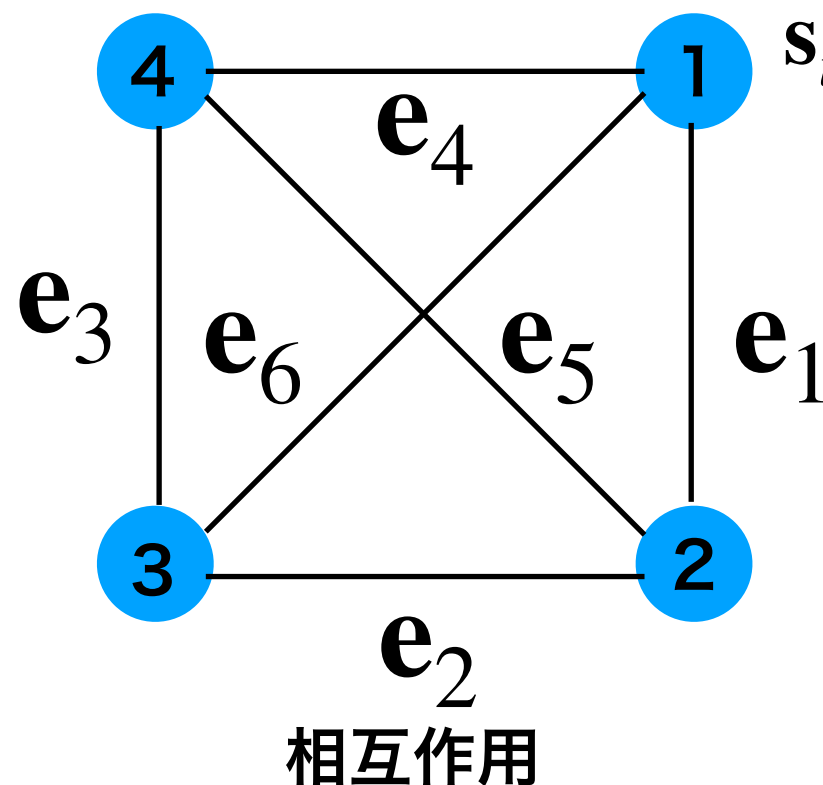
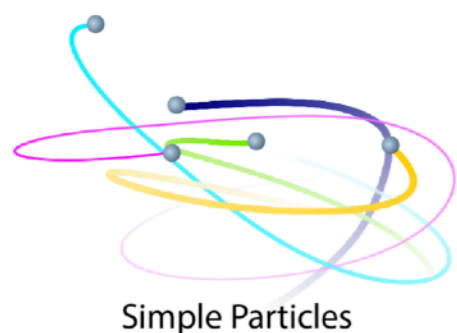


[Arrieta, Alejandro Barredo, et al., Information fusion, 58, 2020.]

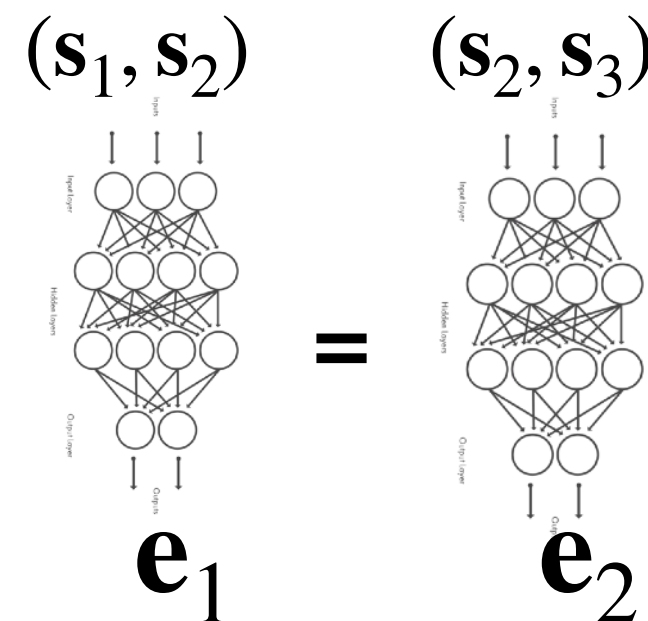
▶ 理学では最終的には初等関数の組み合わせで法則を表現したい?

## ○記号回帰によるDNNのsimplification

[Cranmer, Miles, et al., Neurips2020]



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



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:  $\mathbf{e}'_k = \phi^e(\mathbf{v}_{r_k}, \mathbf{v}_{s_k})$ ,

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

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

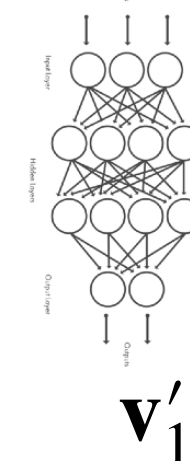
$$\bar{\mathbf{e}}'_i \in \mathbb{R}^{L^{e'}}$$

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

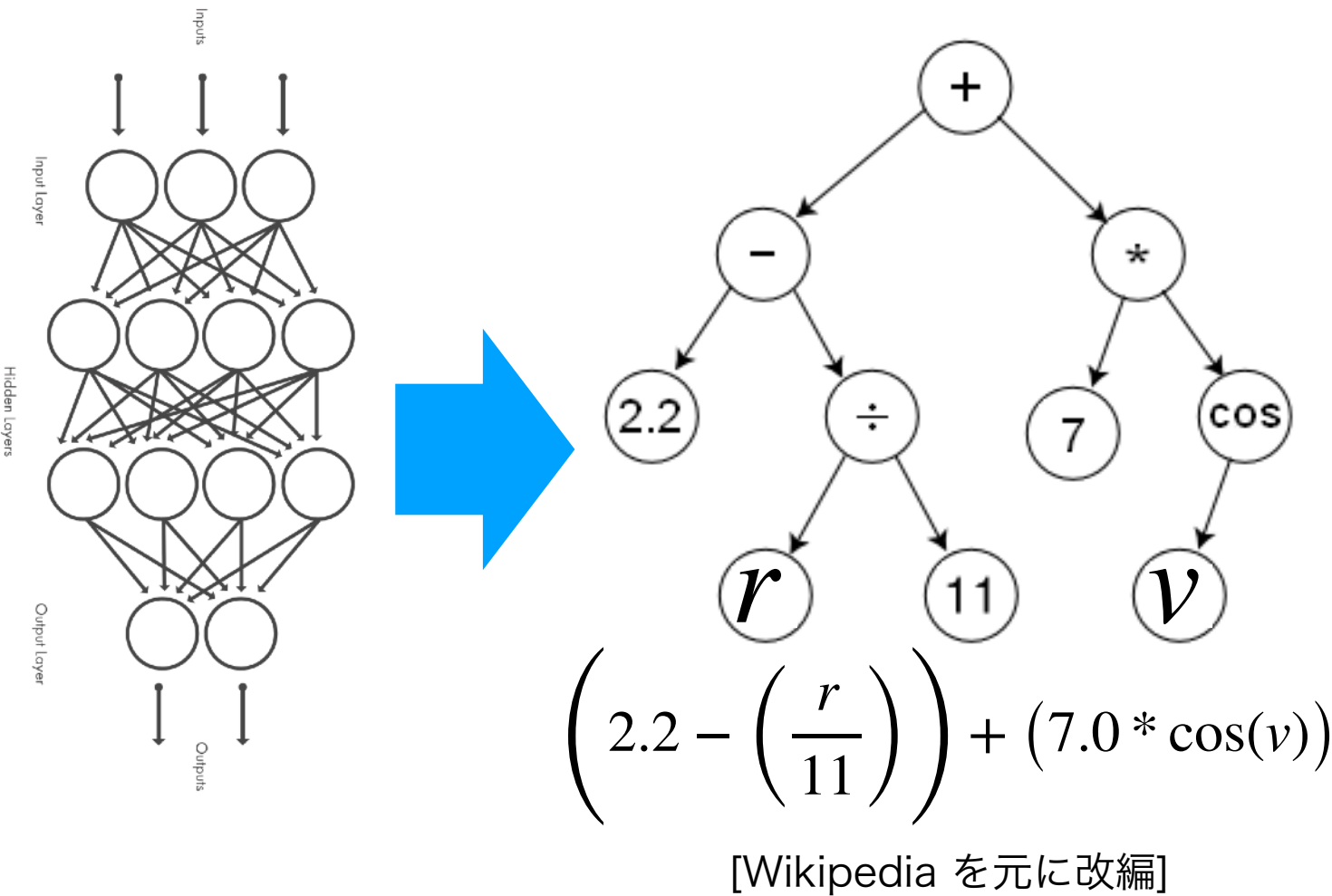
$$\hat{\mathbf{v}}'_i \in \mathbb{R}^{L^{v'}}$$

力

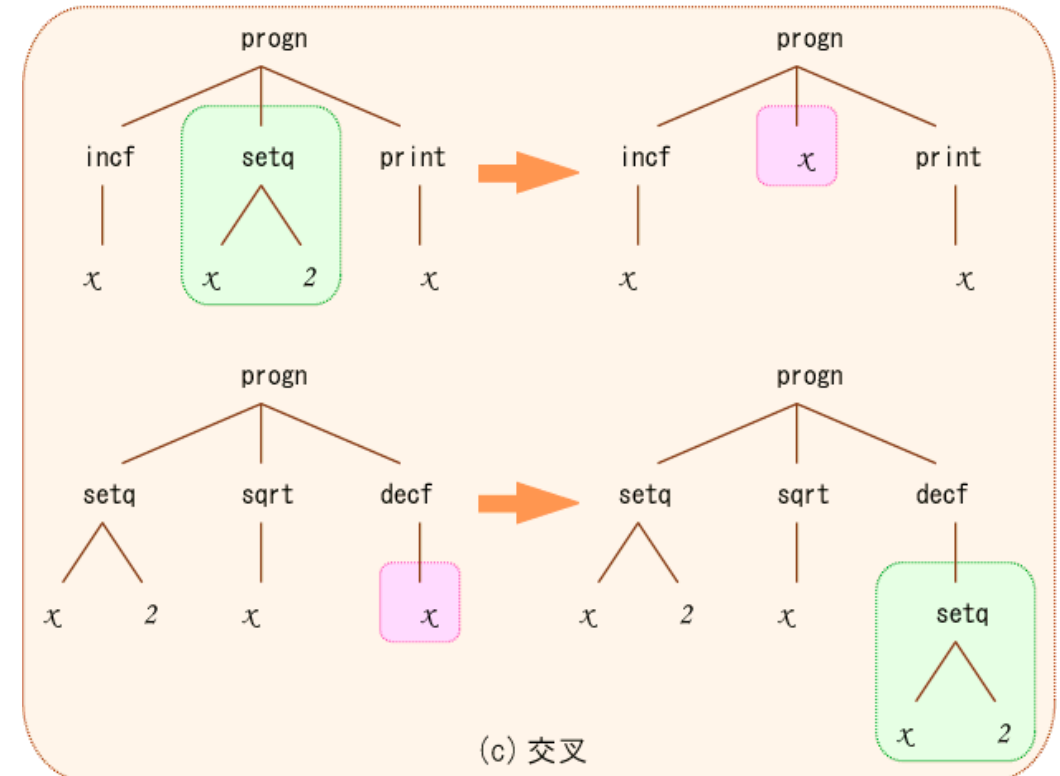
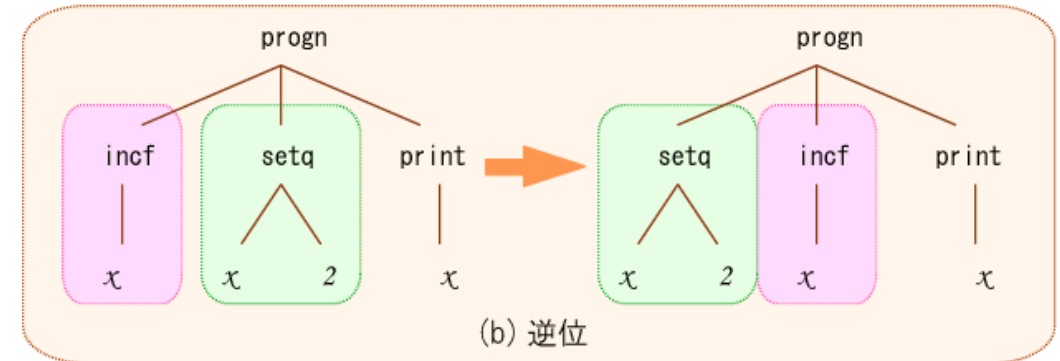
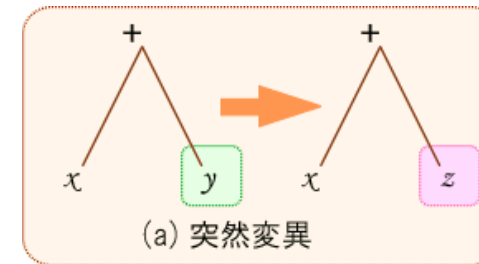
$$\left( \mathbf{s}_1, \sum_k \mathbf{e}_k \right)$$



## ○記号回帰によるDNNのsimplification



## 遺伝的プログラミング



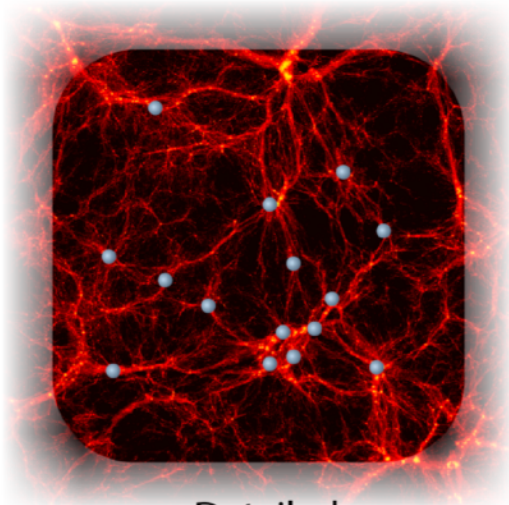
[伊庭研究室 HP より]

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



## ○記号回帰によるDNNのsimplification

[Cranmer, Miles, et al., Neurips2020]



Detailed Dark Matter Simulation

Dark matterが集まったHaloを粒子とみなして、  
その質量、相互距離、速度と  
Haloの中心密度 (に関連する以下の量) の関係則を知りたい

$$\delta = \frac{\rho - \langle \rho \rangle}{\langle \rho \rangle}$$

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

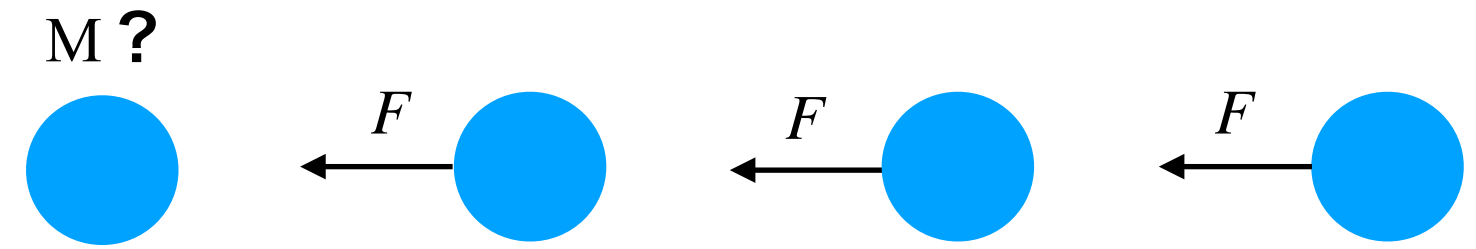
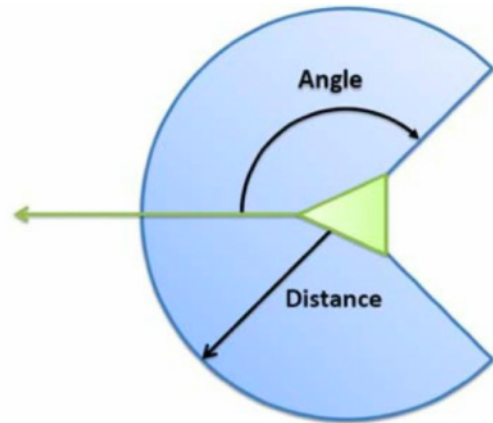
Table 2: A comparison of both known and discovered formulas for dark matter overdensity.  $C_i$  indicates fitted parameters, which are given in the appendix.

▶現象をより良く表す新法則を発見



**[step1]** 有限のデータ・精度で物理モデルを帰納的に推定した場合に、物理モデルが一意に定まるか（不定性がどの程度あるか）の確認

例：視野角ありの相互作用

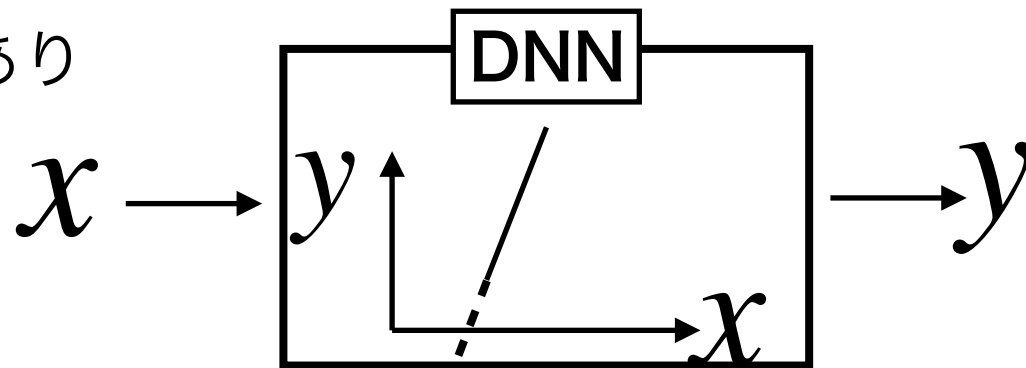


左端粒子のみ大質量 or 視野ありの引力相互作用？

[<https://www.wallacelages.com/boids.html>]

**[step2]** 有限データで学習した機械学習モデルから物理的に妥当な情報を抽出するために、解釈手法に事前知識を導入する

例：値域あり



【物理的制約： $y > 0$ 】

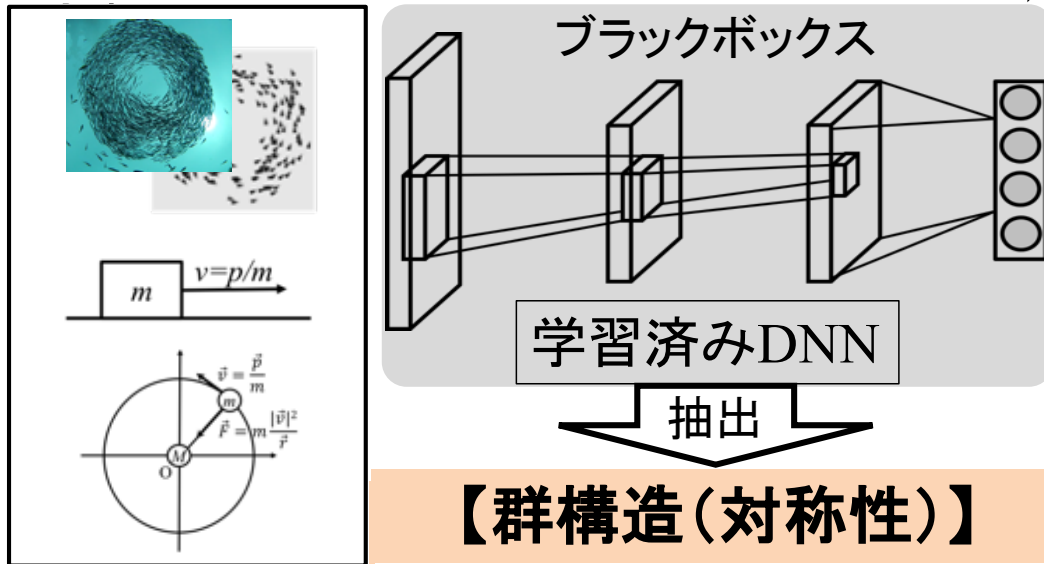
~~線形： $y = ax + b$~~

シグモイド関数： $y = \frac{1}{1 + \exp^{-(ax+b)}}$

**[step3]** 解釈手法を計測装置だと考えて解釈手法のハイパーパラメータなどを探索しながら知識抽出

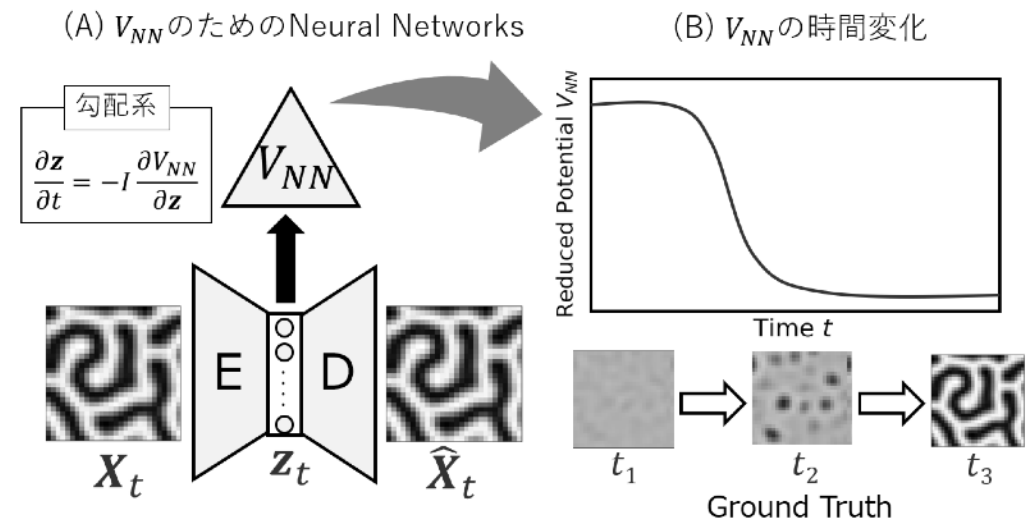
例：後ほど、事例紹介で説明します

## 力学系の対称性・保存則の抽出 (Model simplification)



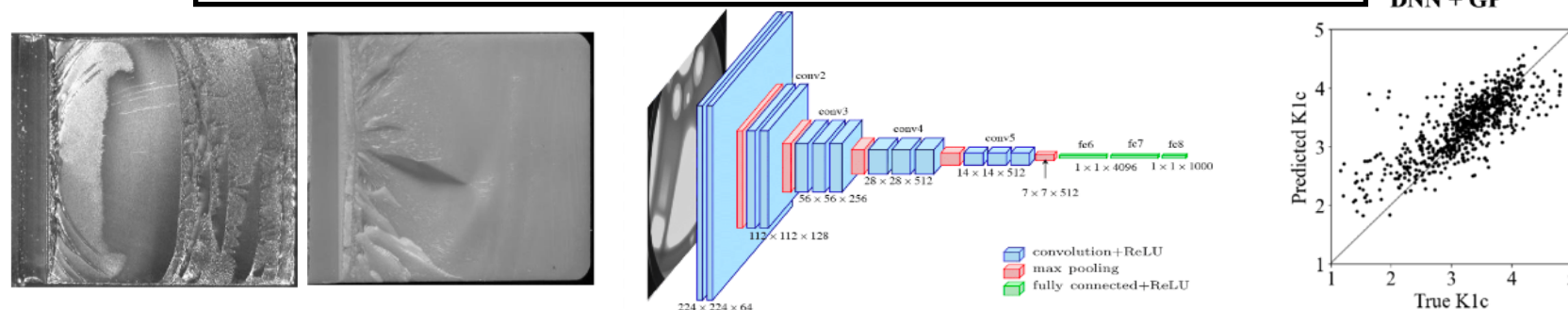
[Y. Mototake, Phys. Rev. E 103, 2021]

## 動画像データからのポテンシャル関数抽出 (Local explanations)



[S. Tsuji, Y. Mototake, et al., in prep]

## 材料パターン画像からの力学特性値の推定 (Explanation by example)



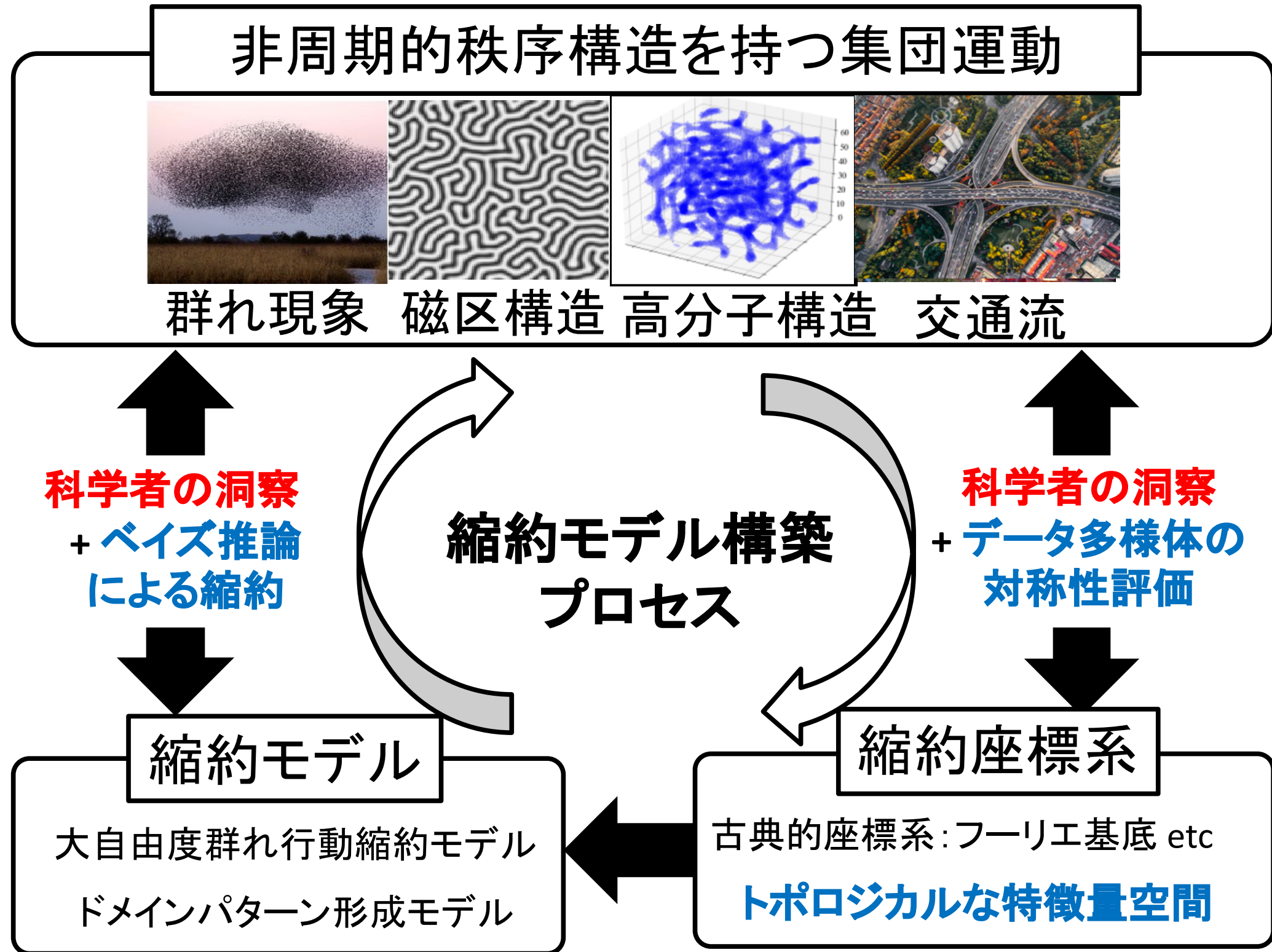
[Y. Mototake, K. Ito, M. Demura, 2022]

## **Interpretable conservation law estimation by deriving the symmetries of dynamics from trained deep neural networks**

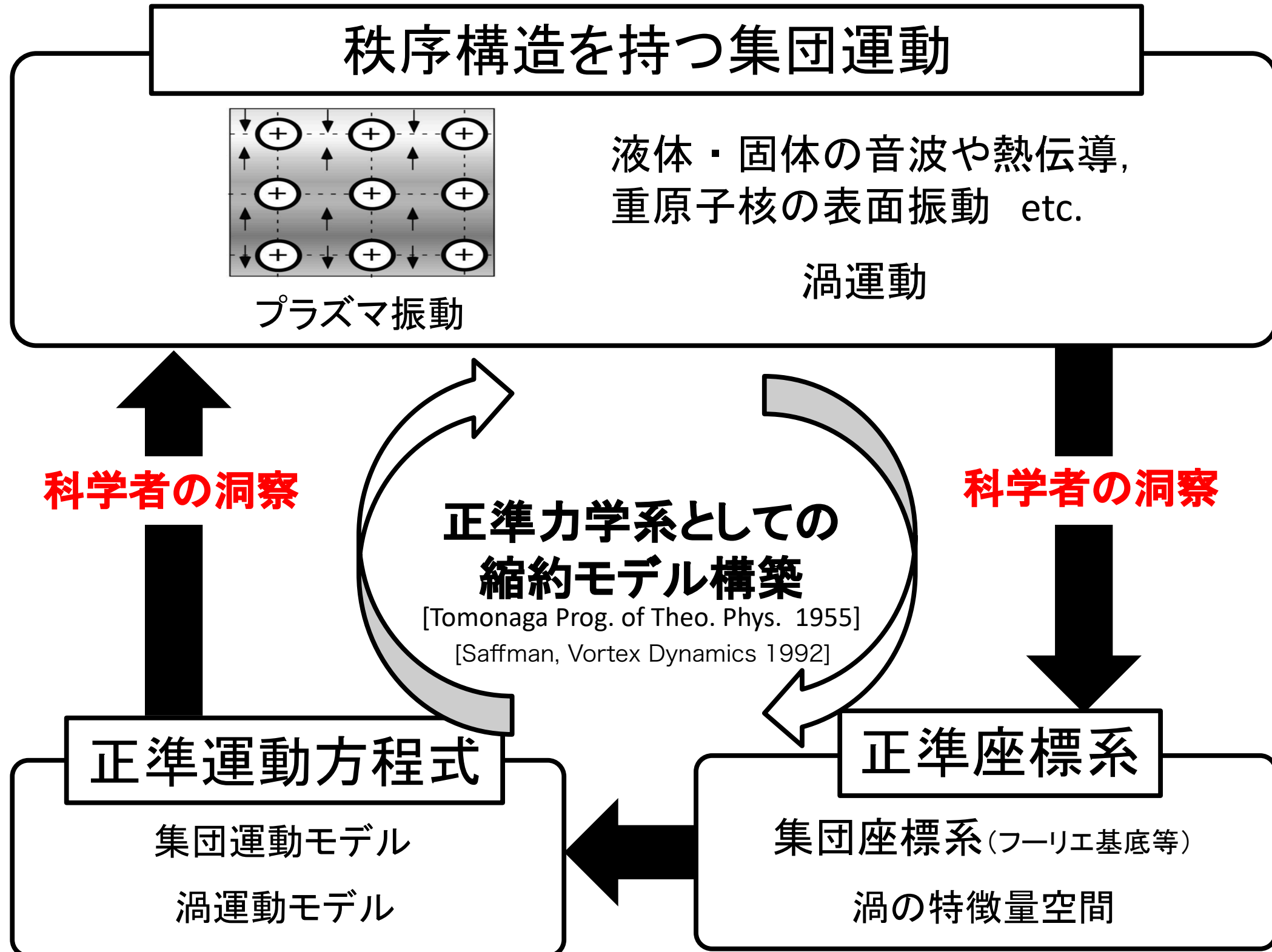
Yoh-ichi Mototake<sup>1</sup>

<sup>1</sup>The Institute of Statistical Mathematics

Phys. Rev. E, 103, 033303, (2021)

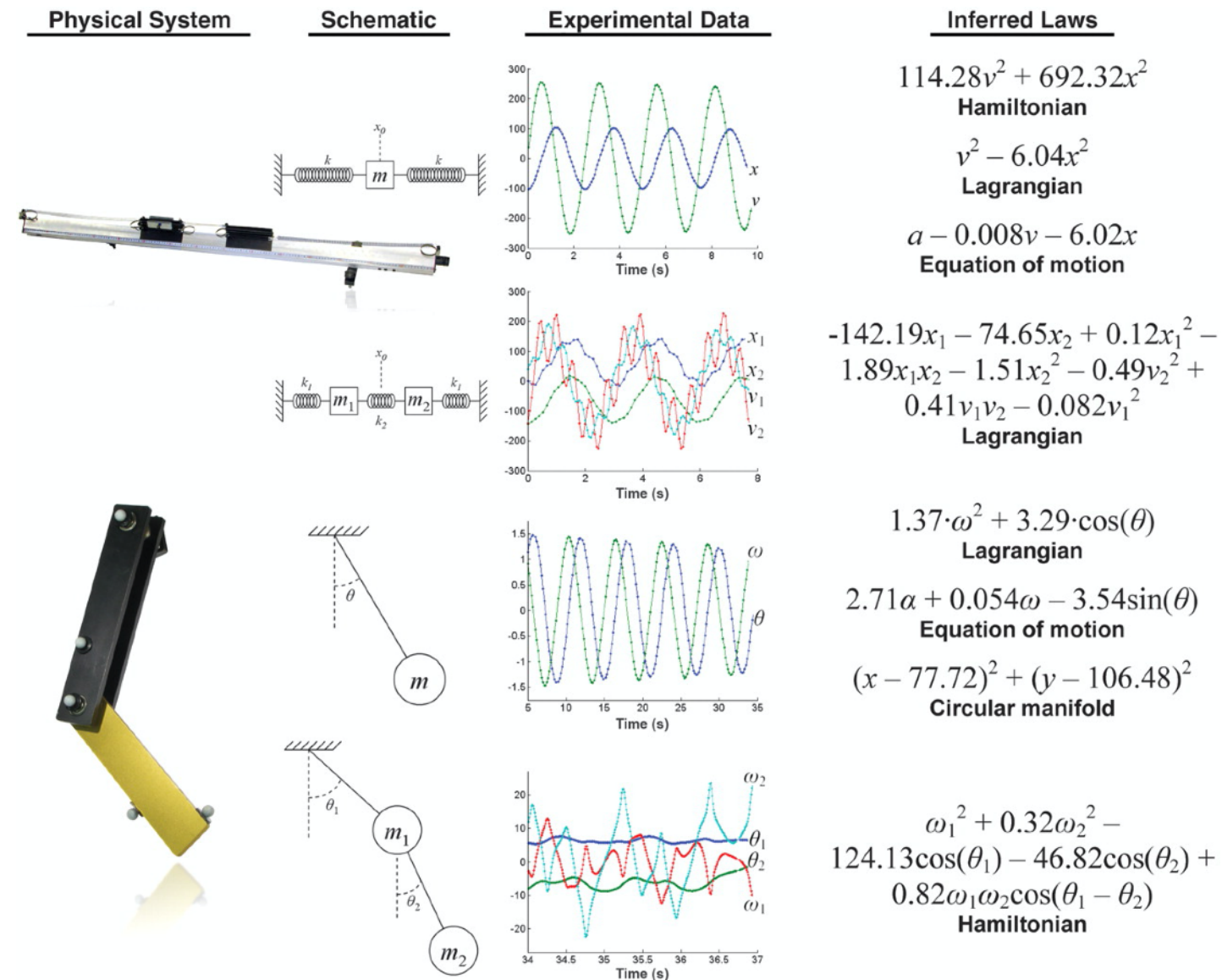
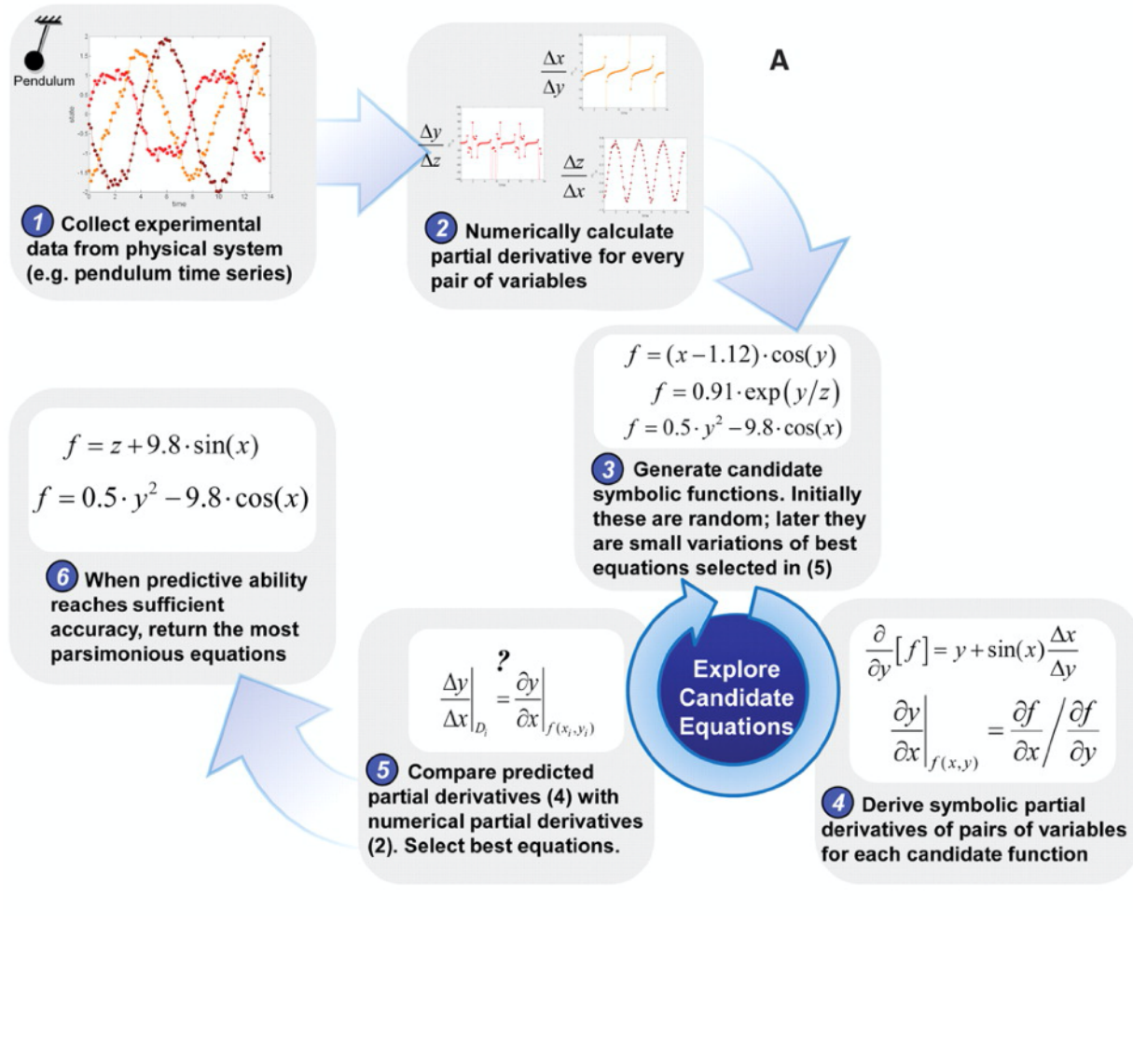


- 集団運動が完全に他の運動と分離されている場合に有効な枠組み [澤田 80]





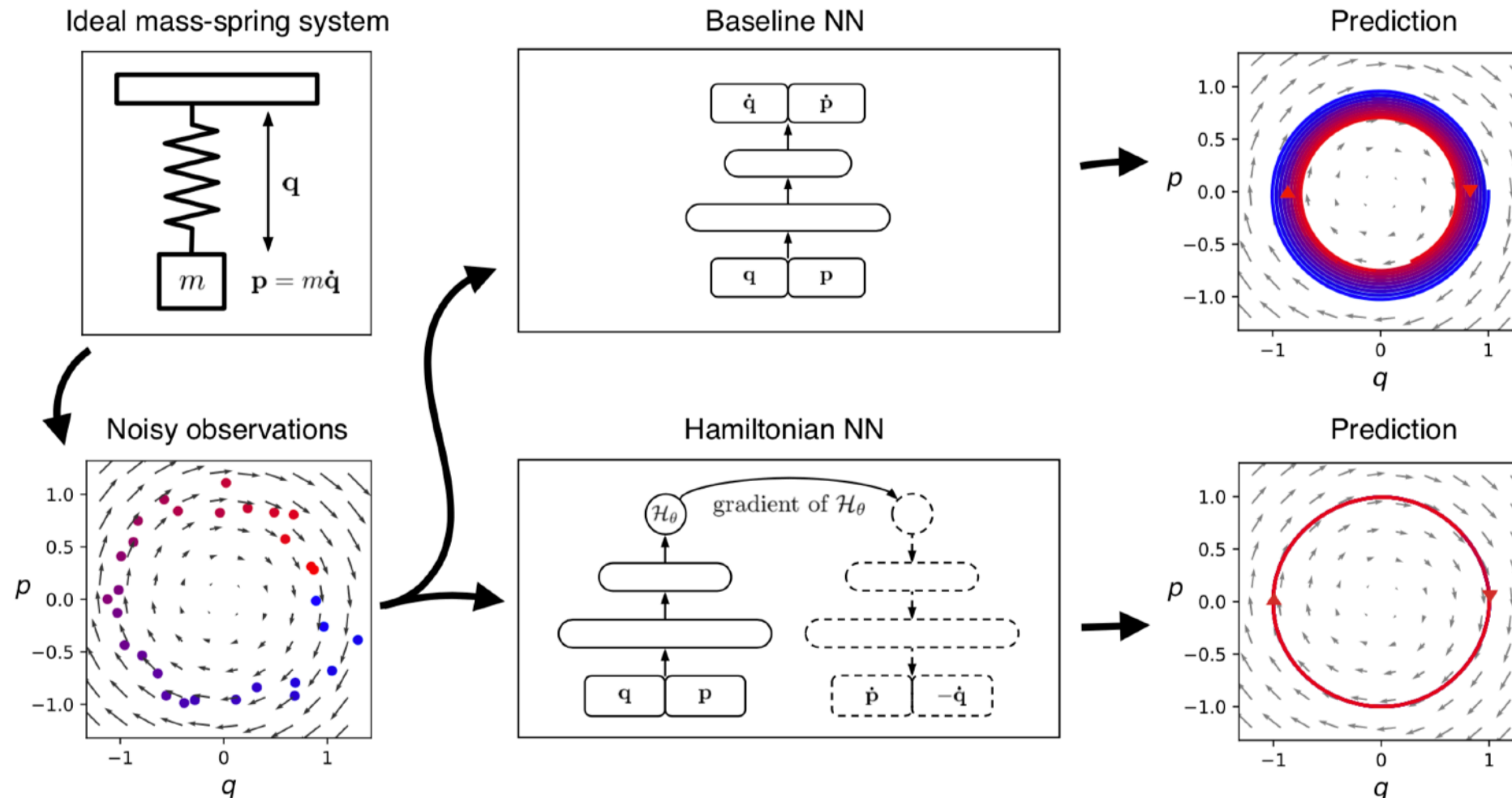
## ● 基底関数の結合によってハミルトニアンを構築



▶ 基底関数が未知の場合には対応が難しい

[Michael Schmidt], and Hod Lipson, Science 2009]

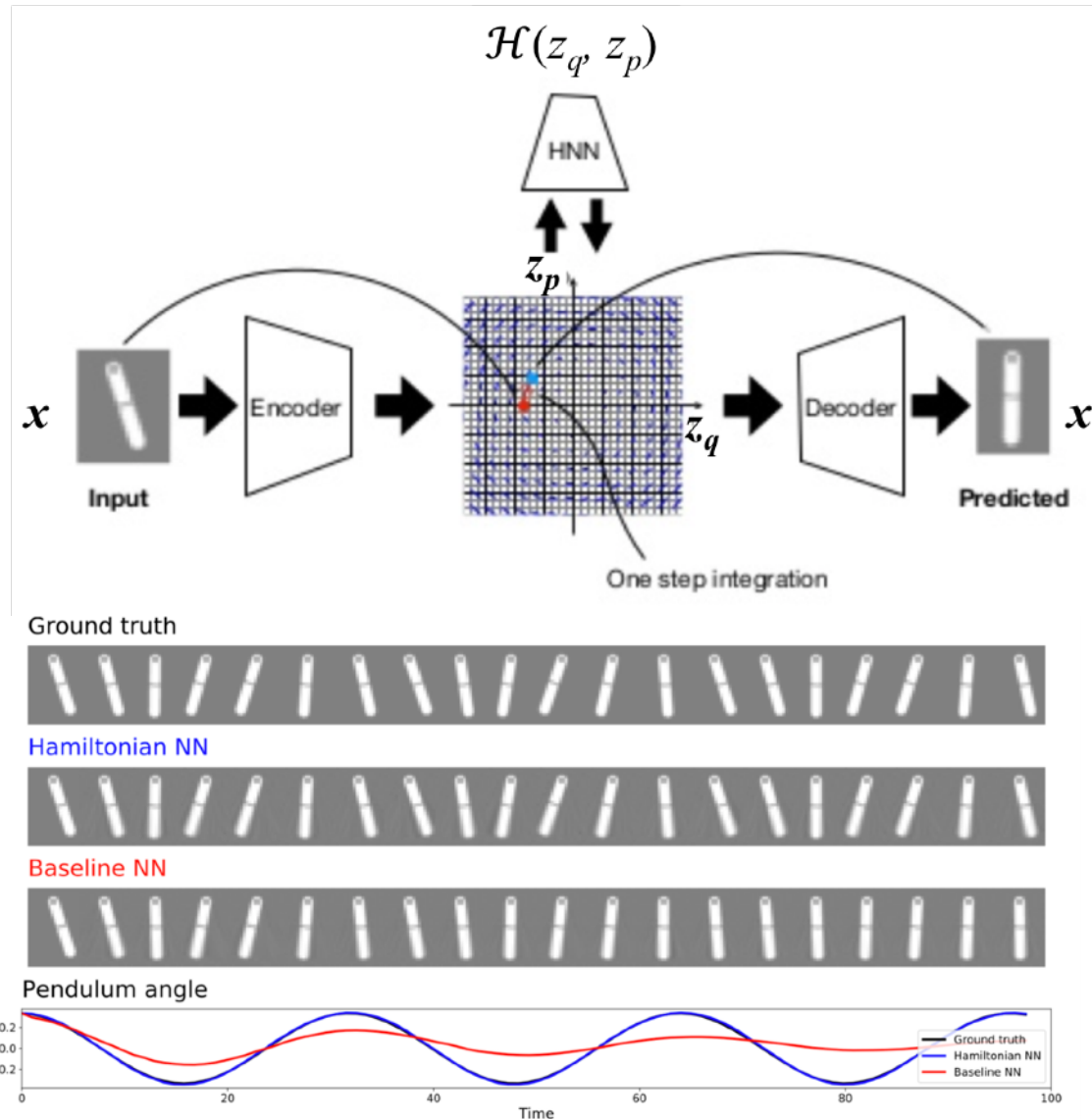
● ハミルトニアン系の制約を導入したDNN



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

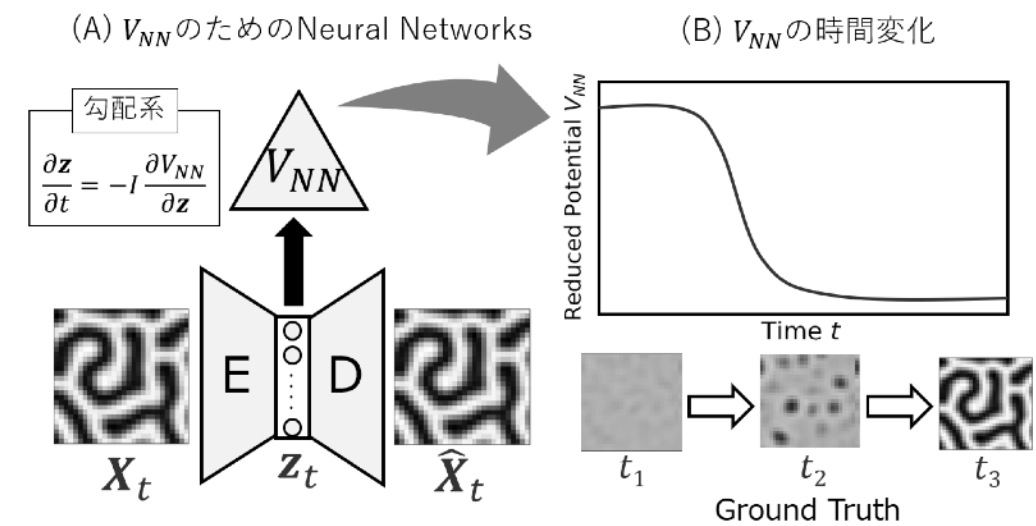
[G. Samuel, M. Dzamba, and J. Yosinski, NeurIPS 2019]

## ● 深層学習による縮約モデルの構築



[G. Samuel, M. Dzamba, and J. Yosinski, NeurIPS 2019]

## 動画データからのポテンシャル関数抽出

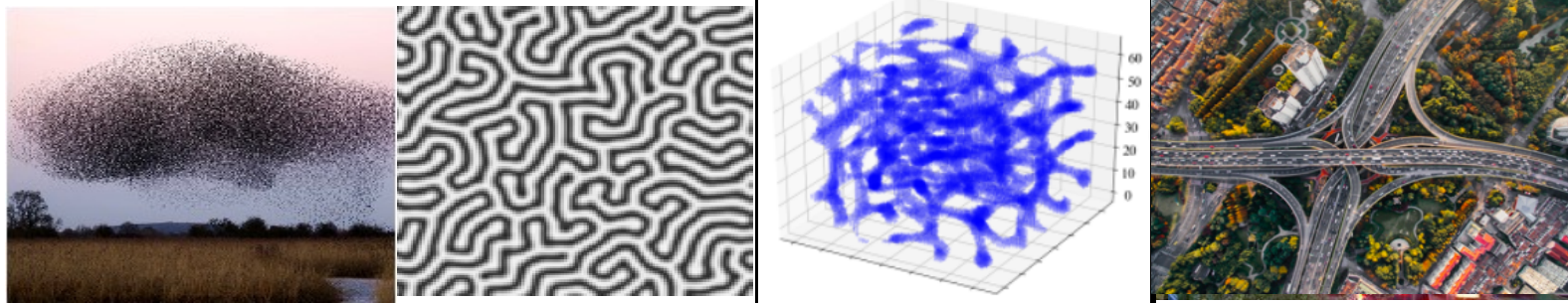


[S. Tsuji, Y. Mototake, et al., in prep]

▶ ハミルトニアンは基本的にブラックボックス



## 非周期的秩序構造を持つ集団運動



群れ現象 磁区構造 高分子構造 交通流

科学者の洞察  
+ ベイズ推論  
による縮約

縮約モデル構築  
プロセス

科学者の洞察  
+ データ多様体の  
対称性評価

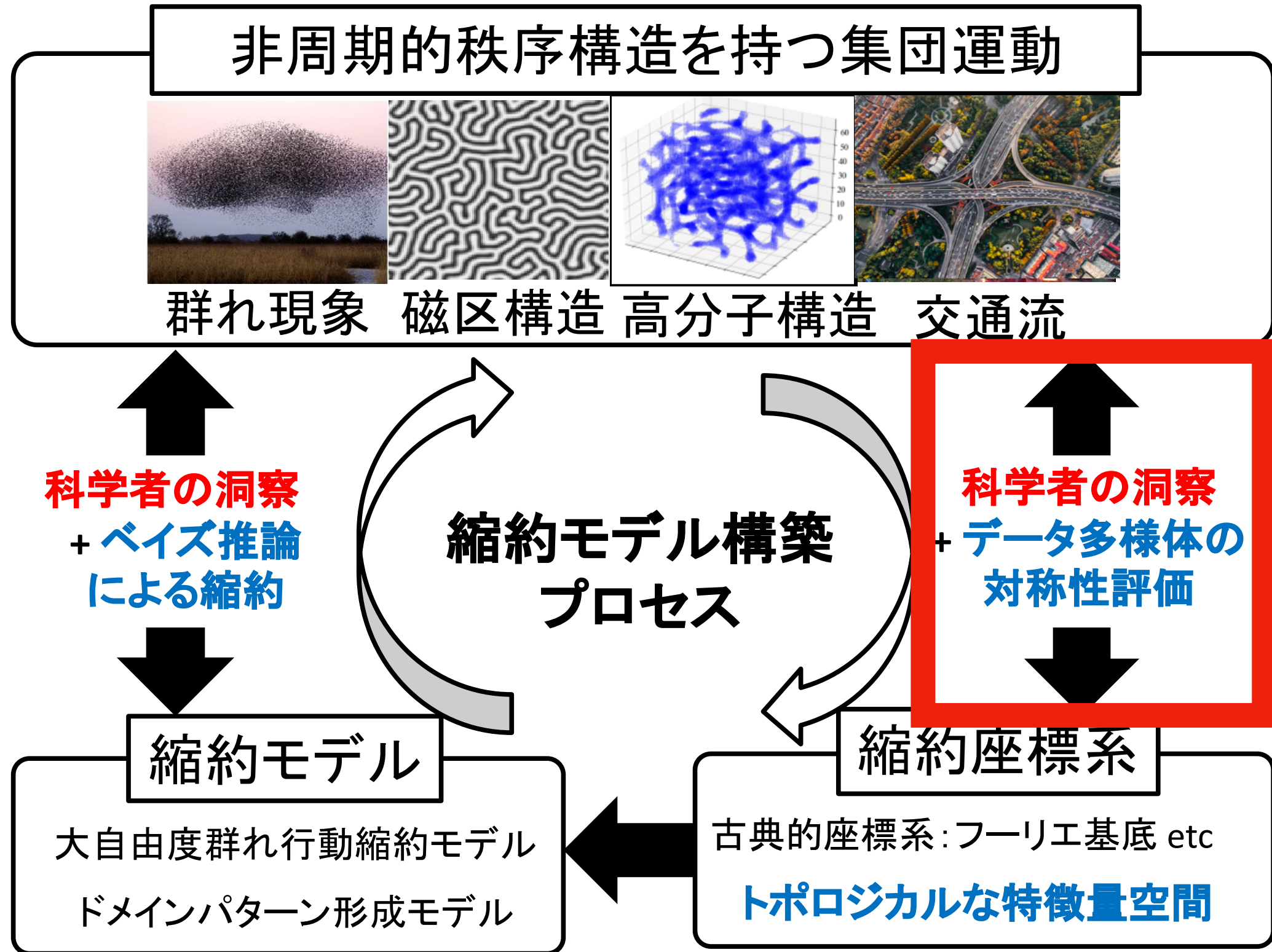
縮約モデル

大自由度群れ行動縮約モデル  
ドメインパターン形成モデル

縮約座標系

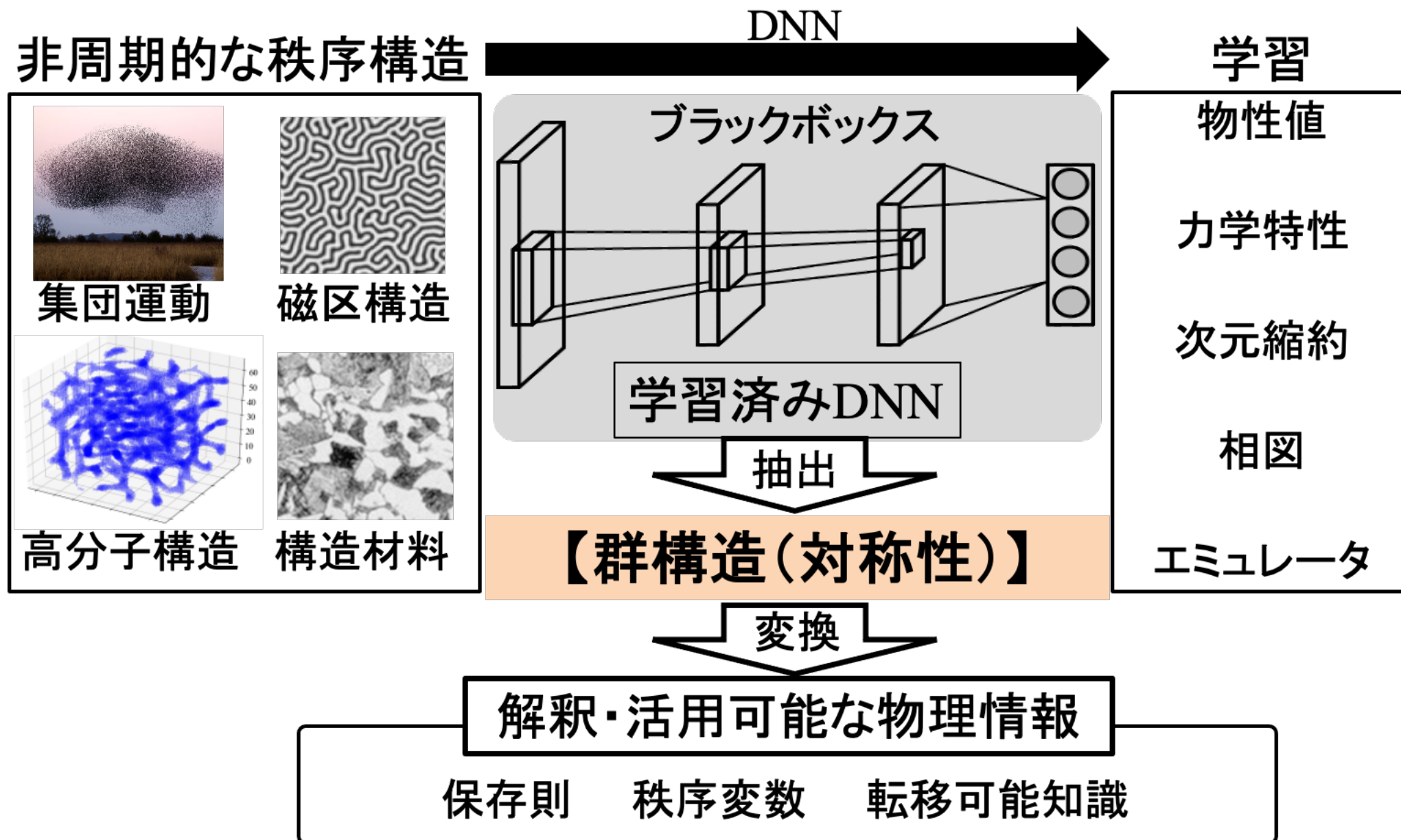
古典的座標系: フーリエ基底 etc  
トポロジカルな特徴量空間

ブラックボックスのハミルトニアン推定



目的：縮約座標系を探索する物理学者に、その座標系を評価できる情報を提示する





力学系を学習済みのDNNから、  
DNNがモデル化した物理モデル（座標系+ハミルトニアン）  
における保存推定を試みる

[Y. Mototake, Phys. Rev. E, 103, 033303, 2021]

- ▶ 提案するフレームワークの目的は、ディープラーニングで物理データを解析することではなく、学習したDNNから解釈可能な物理情報を抽出することである。
- ▶ 提案するフレームワークは、Noetherの定理と効率的なサンプリング手法により、学習済みDNNからダイナミクスの対称性を抽出することで、保存則を推論するものです。
- ▶ 提案する枠組みは、時系列データセットの多様体構造とNoetherの定理の必要条件との関係を導き出すことによって開発したものである。

同時期に類似研究がいくつか出版（その前に出版でき引用してもらえました）：

[Ziming Liu and Max Tegmark, Phys. Rev. Lett. 126 180604 (2021)]

[Seungwoong Ha and Hawoong Jeong, Phys. Rev. Research 3 L042035 (2021)]

[Han Zhang, Huawei Fan, Liang Wang, and Xingang Wang, Phys. Rev. E 104 024205 (2021)]

## 前提確認 ([step1]に対応)

1. ネーターの定理と時系列データ多様体の関係を検討
2. DNNによるデータ多様体のモデル化について検討

## 手法

3. DNNから系の対称性を抽出する手法の説明
4. 抽出された対称性から保存則を推定する手法の説明

## 前提情報 ([step1]に対応)

1. ネーターの定理と時系列データ多様体の関係を検討
2. DNNによるデータ多様体のモデル化について検討

## 手法

3. DNNから系の対称性を抽出する手法の説明
4. 抽出された対称性から保存則を推定する手法の説明



## ネーターの定理：

ハミルトン系の持つ連続対称性と保存則を結びつける定理 [Noether 1918].

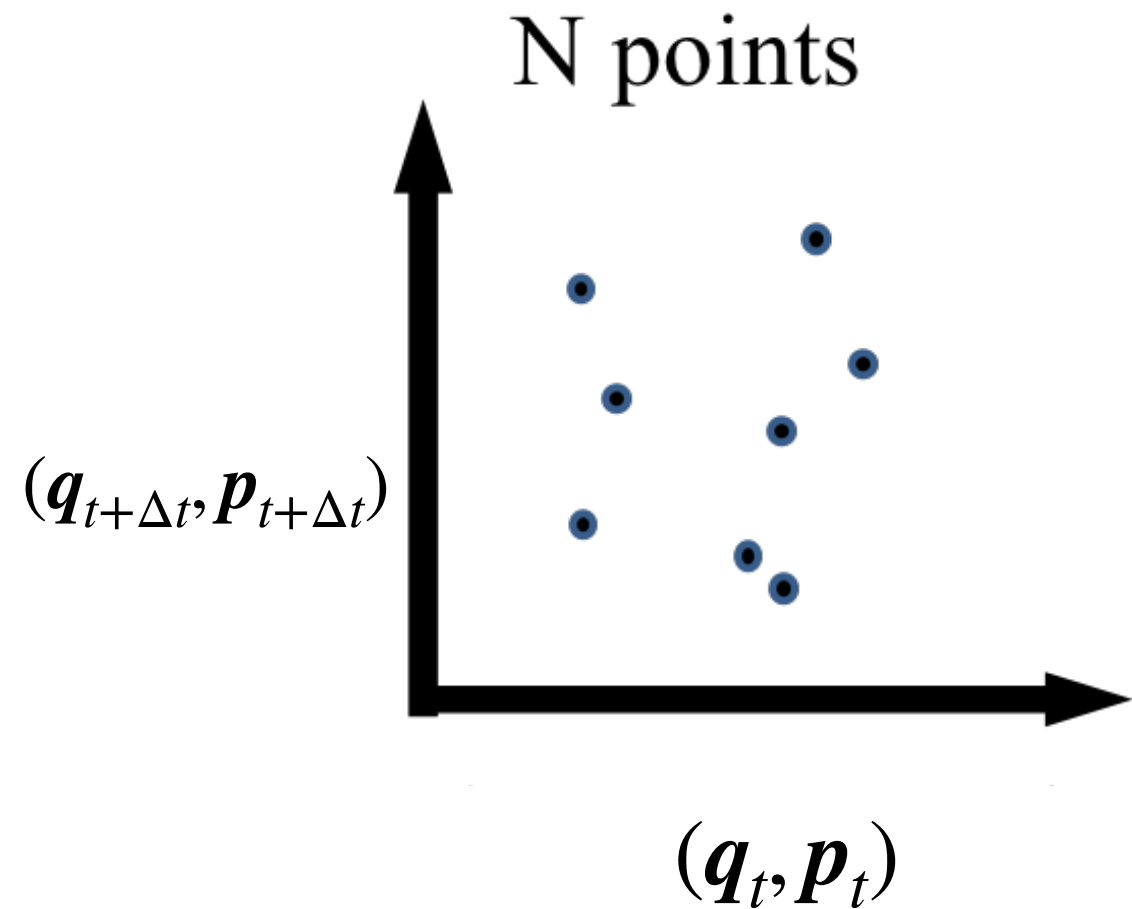
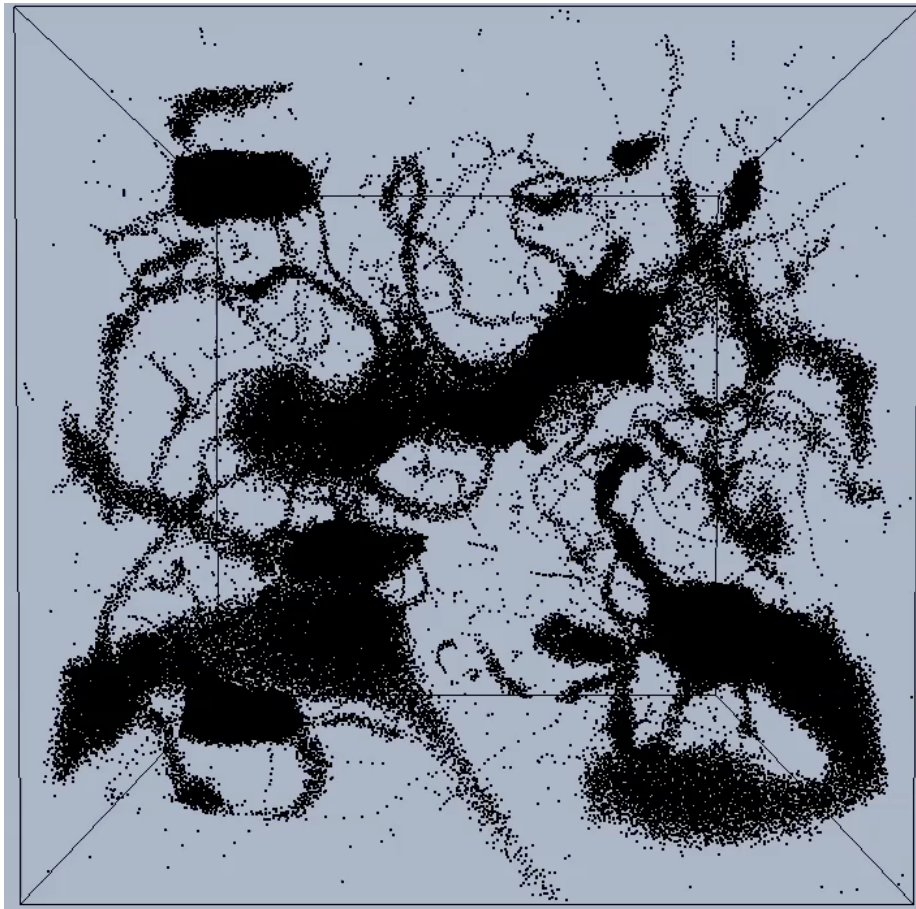
ハミルトニアン $H(\mathbf{q}, \mathbf{p})$ と正準運動方程式 $\frac{\partial H(\mathbf{q}, \mathbf{p})}{\partial q_i} = -\dot{p}_i$ ,  $\frac{\partial H(\mathbf{q}, \mathbf{p})}{\partial p_i} = \dot{q}_i$  が,

無限小変換 $(q'_i, p'_i) = (q_i + \delta q_{ij}, p_i + \delta p_{ij})$ に対して不変であるとするとき、保存量 $G_j$ との間に以下の関係式が成立する。

$$(\delta q_{ij}, \delta p_{ij}) = \left( \frac{\partial G_j}{\partial p_i}, -\frac{\partial G_j}{\partial q_i} \right)$$

本研究における力学系時系列データの定義

$$D := \left\{ \mathbf{q}_{t_i}^i, \mathbf{p}_{t_i}^i, \mathbf{q}_{t_i+\Delta t}^i, \mathbf{p}_{t_i+\Delta t}^i \right\}_{i=1}^N$$



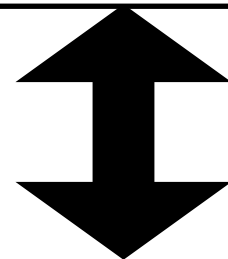
## ハミルトン系の対称性

ハミルトニアンの不変性： $H'(q, p) \equiv H(q, p)$

$$H'(Q, P) := H(q(Q, P), p(Q, P))$$

正準運動方程式：
$$\begin{cases} \mathbf{q}_{T+\Delta T} = \frac{\partial H(\mathbf{q}_T, \mathbf{p}_T)}{\partial \mathbf{p}_T} \Delta T + \mathbf{q}_T \\ \mathbf{p}_{T+\Delta T} = -\frac{\partial H(\mathbf{q}_T, \mathbf{p}_T)}{\partial \mathbf{q}_T} \Delta T + \mathbf{p}_T \end{cases}$$
 の不変性

$$D := \left\{ \mathbf{q}_{t_i}^i, \mathbf{p}_{t_i}^i, \mathbf{q}_{t_i+\Delta t}^i, \mathbf{p}_{t_i+\Delta t}^i \right\}_{i=1}^N$$

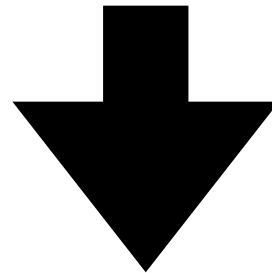


## 力学系時系列データ多様体の対称性

$$\begin{aligned} & \forall E, \left\{ \mathbf{q}_{t+\Delta t}, \mathbf{p}_{t+\Delta t}, \mathbf{q}_t, \mathbf{p}_t \mid H(\mathbf{q}_t, \mathbf{p}_t) = E, \mathbf{p}_{t+\Delta t} = \mathbf{p}_t - \frac{\partial H(\mathbf{q}_t, \mathbf{p}_t)}{\partial \mathbf{q}_t} \Delta t, \mathbf{q}_{t+\Delta t} = \mathbf{q}_t + \frac{\partial H(\mathbf{q}_t, \mathbf{p}_t)}{\partial \mathbf{p}_t} \Delta t \right\} \\ & = \left\{ \mathbf{Q}_{T+\Delta T}, \mathbf{P}_{T+\Delta T}, \mathbf{Q}_T, \mathbf{P}_T \mid H(\mathbf{q}_t, \mathbf{p}_t) = E, \mathbf{p}_{t+\Delta t} = \mathbf{p}_t - \frac{\partial H(\mathbf{q}_t, \mathbf{p}_t)}{\partial \mathbf{q}_t} \Delta t, \mathbf{q}_{t+\Delta t} = \mathbf{q}_t + \frac{\partial H(\mathbf{q}_t, \mathbf{p}_t)}{\partial \mathbf{p}_t} \Delta t \right\} \end{aligned}$$

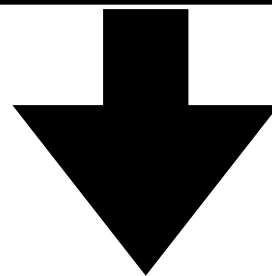
問題の困難さの緩和：全エネルギーという条件の緩和

$$\begin{aligned} & \forall E, \left\{ \mathbf{q}_{t+\Delta t}, \mathbf{P}_{t+\Delta t}, \mathbf{q}_t, \mathbf{p}_t \mid H(\mathbf{q}_t, \mathbf{p}_t) = E, \mathbf{p}_{t+\Delta t} = \mathbf{p}_t - \frac{\partial H(\mathbf{q}_t, \mathbf{p}_t)}{\partial \mathbf{q}_t}, \mathbf{q}_{t+\Delta t} = \mathbf{q}_t + \frac{\partial H(\mathbf{q}_t, \mathbf{p}_t)}{\partial \mathbf{p}_t} \right\} \\ & = \left\{ \mathbf{Q}_{T+\Delta T}, \mathbf{P}_{T+\Delta T}, \mathbf{Q}_T, \mathbf{P}_T \mid H(\mathbf{q}_t, \mathbf{p}_t) = E, \mathbf{p}_{t+\Delta t} = \mathbf{p}_t - \frac{\partial H(\mathbf{q}_t, \mathbf{p}_t)}{\partial \mathbf{q}_t}, \mathbf{q}_{t+\Delta t} = \mathbf{q}_t + \frac{\partial H(\mathbf{q}_t, \mathbf{p}_t)}{\partial \mathbf{p}_t} \right\} \end{aligned}$$



$$\{\mathbf{Q}_i(\mathbf{q}, \mathbf{p}), \mathbf{P}_i(\mathbf{q}, \mathbf{p}) \mid \forall i \in \Lambda_E, \text{ satisfy Eq. (32)}\} = \bigcap_{i \in \Lambda_E} \{\mathbf{Q}_i(\mathbf{q}, \mathbf{p}), \mathbf{P}_i(\mathbf{q}, \mathbf{p}) \mid \text{ satisfy Eq. (32)}\}.$$

あるエネルギーで等式が成り立つ条件が必要条件



$$S_i := \left\{ \mathbf{q}_{t+\Delta t}, \mathbf{P}_{t+\Delta t}, \mathbf{q}_t, \mathbf{p}_t \mid H(\mathbf{q}_t, \mathbf{p}_t) = E_i, \mathbf{p}_{t+\Delta t} = \mathbf{p}_t - \frac{\partial H(\mathbf{q}_t, \mathbf{p}_t)}{\partial \mathbf{q}_t}, \mathbf{q}_{t+\Delta t} = \mathbf{q}_t + \frac{\partial H(\mathbf{q}_t, \mathbf{p}_t)}{\partial \mathbf{p}_t} \right\}$$

▶ ある特定エネルギー  $E_i$  での時系列データ多様体  $S_i$  の対称性を調べるのが目標になった！！

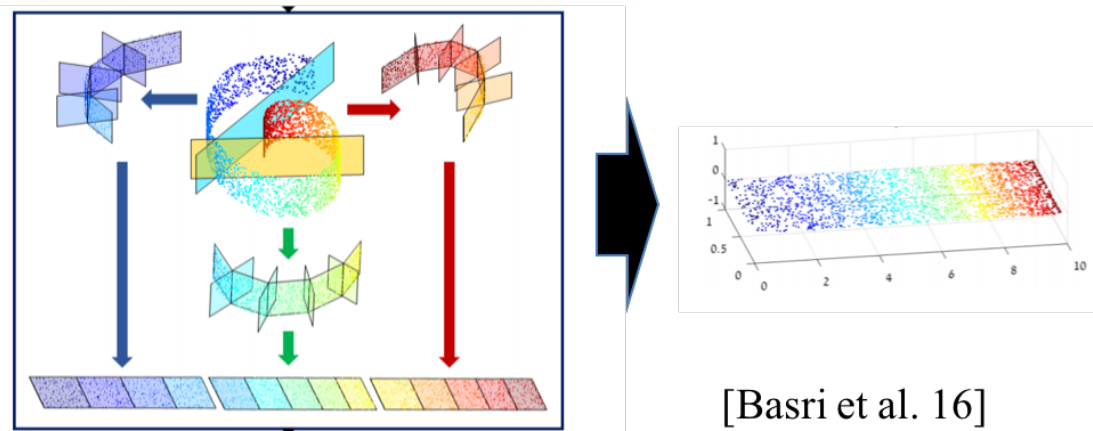
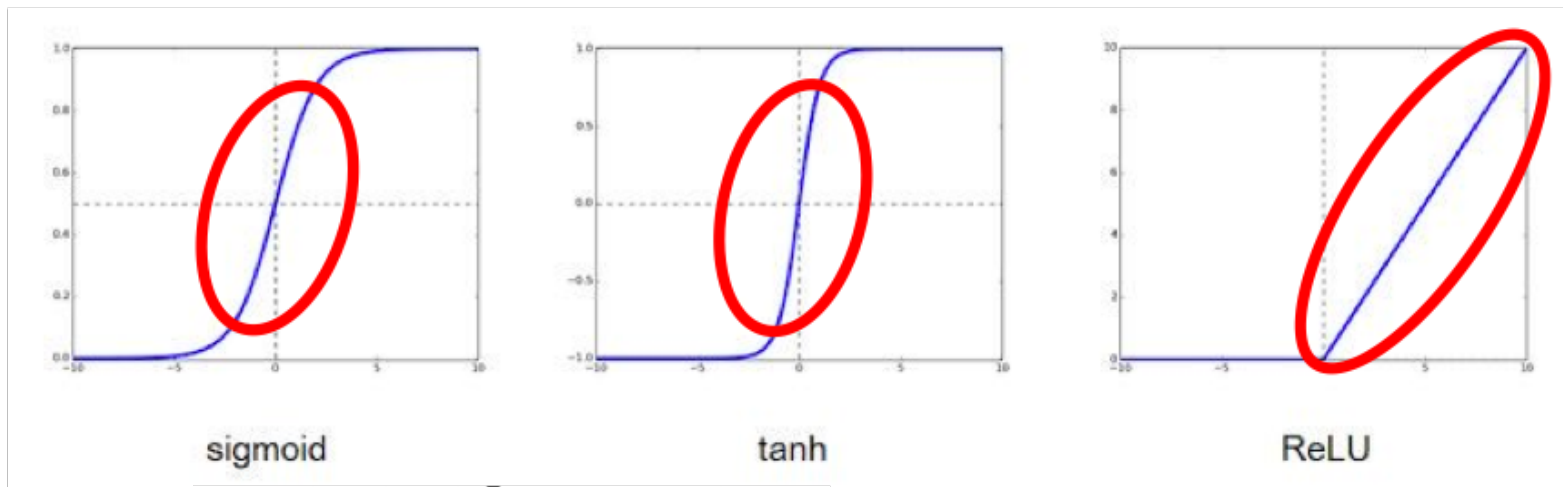


## ニューラルネットとデータ多様体

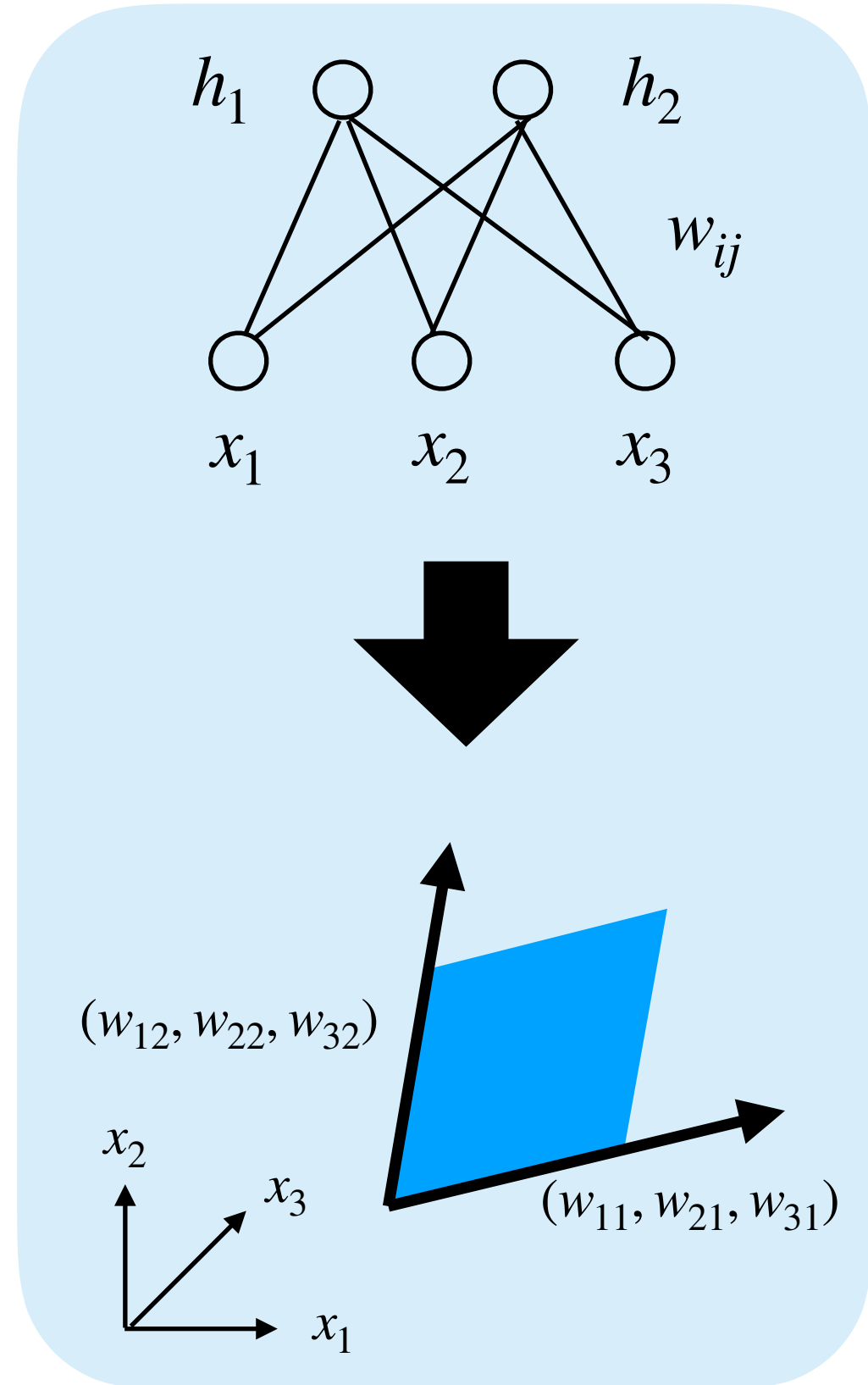
$$\mathbf{h} = (h_1, h_2, \dots, h_{d_h})$$

$$\varphi(\mathbf{w}^{\text{in}} \mathbf{x}) = (\varphi_1, \varphi_2, \dots, \varphi_{d_h}), \varphi_j = \varphi \left[ \sum_i^{d_{\text{in}}} (w_{ij}^{\text{in}} x_i) \right]$$

活性化関数： $\varphi(x)$



▶ ニューラルネットは、区分的な超平面の張り合わせで多様体をモデル化可能。

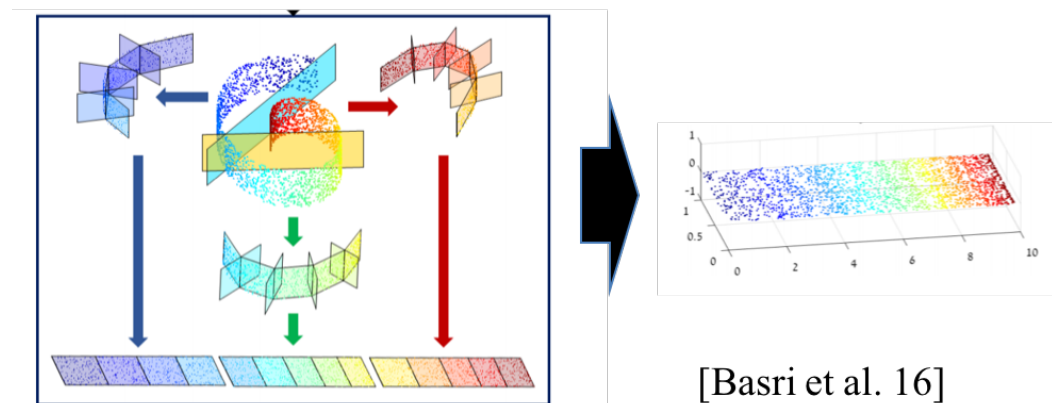


- 時系列データ多様体 $S_i$ の座標変換に対する対称性をみることで、保存則がわかる.

$$S_i := \left\{ \mathbf{q}_{t+\Delta t}, \mathbf{p}_{t+\Delta t}, \mathbf{q}_t, \mathbf{p}_t \mid H(\mathbf{q}_t, \mathbf{p}_t) = E_i, \mathbf{p}_{t+\Delta t} = \mathbf{p}_t - \frac{\partial H(\mathbf{q}_t, \mathbf{p}_t)}{\partial \mathbf{q}_t} \Delta t, \mathbf{q}_{t+\Delta t} = \mathbf{q}_t + \frac{\partial H(\mathbf{q}_t, \mathbf{p}_t)}{\partial \mathbf{p}_t} \Delta t \right\}$$

- DNNは多様体を超平面の貼り合わせのような形で表現できる.

▶ 学習済みDNNは $S_i$ をモデル化している？

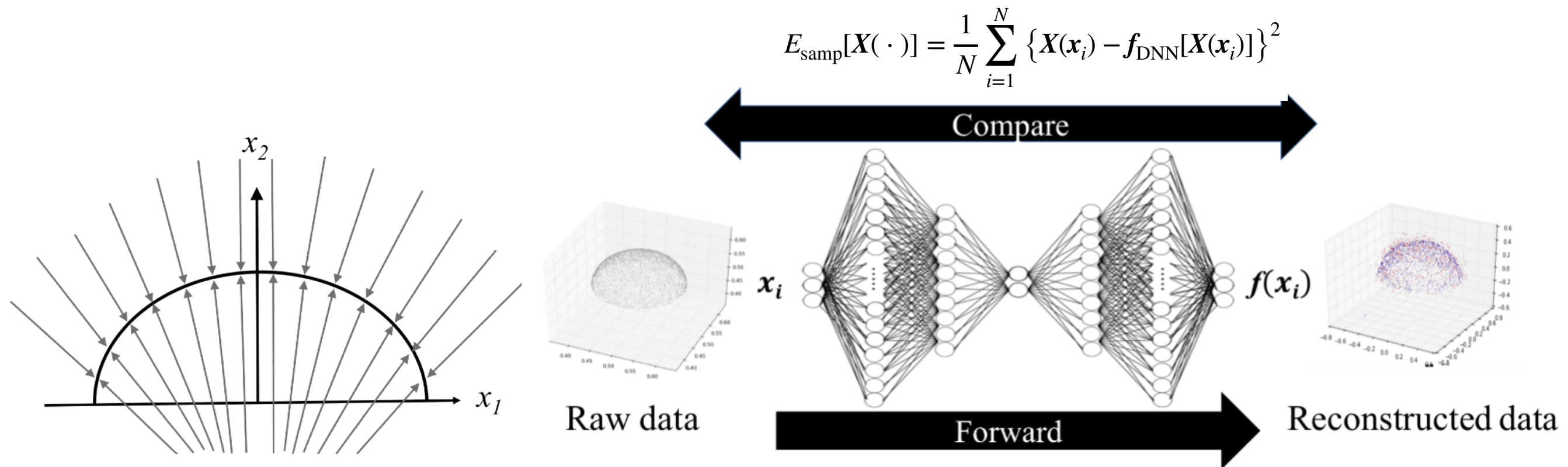


## 前提確認 ([step1]に対応)

1. ネーターの定理と時系列データ多様体の関係を検討
2. DNNによるデータ多様体のモデル化について検討

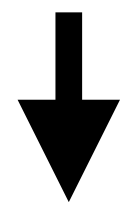
## 手法

3. DNNから系の対称性を抽出する手法の説明
4. 抽出された対称性から保存則を推定する手法の説明



$$\left\{ \mathbf{Q}(\cdot, \cdot), \mathbf{P}(\cdot, \cdot) \mid \arg \min_{\mathbf{Q}(\cdot, \cdot), \mathbf{P}(\cdot, \cdot)} E_{\text{samp}} [\mathbf{Q}(\cdot, \cdot), \mathbf{P}(\cdot, \cdot)] \right\}$$

を満たす変換 $\mathbf{Q}(\cdot, \cdot), \mathbf{P}(\cdot, \cdot)$ の集合として対称変換の集合が得られる

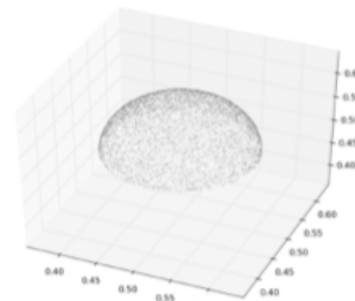
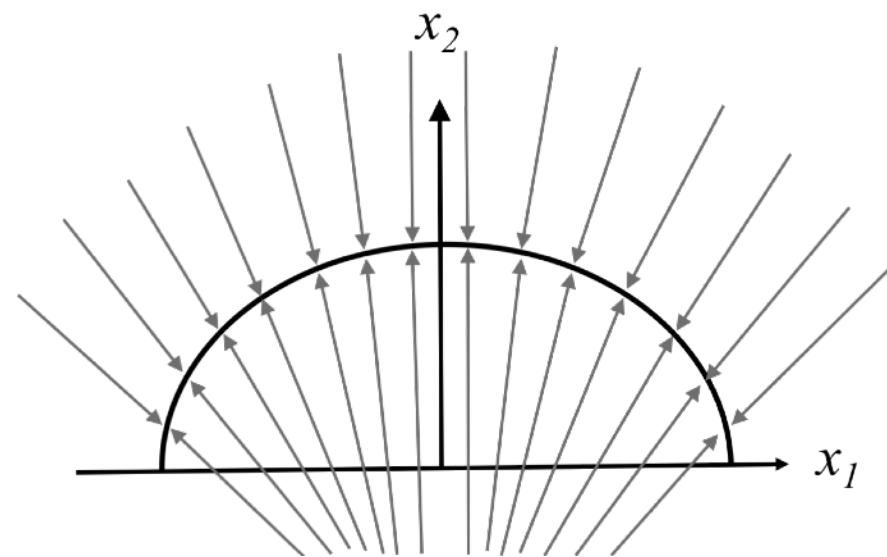
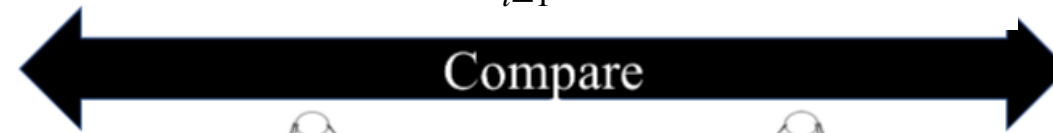


**サンプリング問題に置き換え**

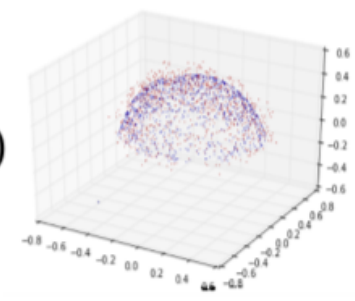
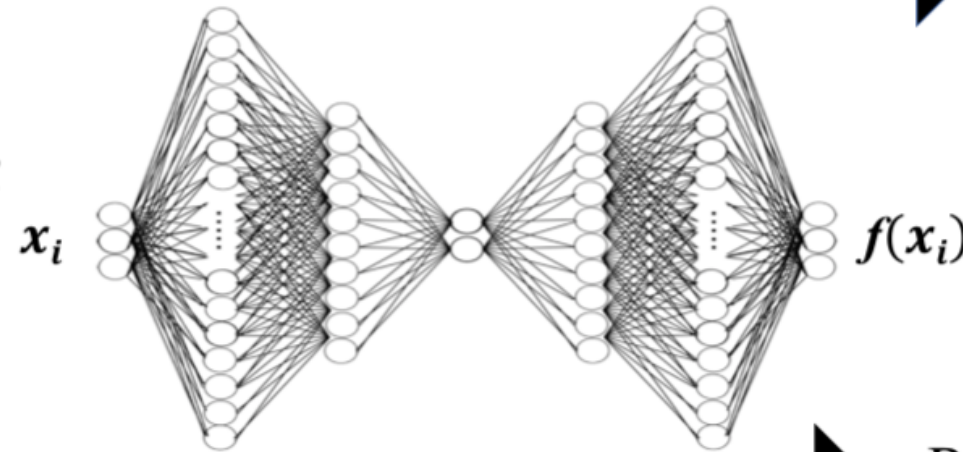
$$P(\mathbf{Q}(\cdot, \cdot), \mathbf{P}(\cdot, \cdot)) \sim \frac{1}{Z} \exp \left\{ -\frac{N}{2\sigma^2} E_{\text{samp}} [\mathbf{Q}(\cdot, \cdot), \mathbf{P}(\cdot, \cdot)] \right\}$$



$$E_{\text{samp}}[\mathbf{X}(\cdot)] = \frac{1}{N} \sum_{i=1}^N \{ \mathbf{X}(\mathbf{x}_i) - f_{\text{DNN}}[\mathbf{X}(\mathbf{x}_i)] \}^2$$



Raw data



Reconstructed data

$$P(\mathbf{Q}(\cdot, \cdot), \mathbf{P}(\cdot, \cdot)) \sim \frac{1}{Z} \exp \left\{ -\frac{N}{2\sigma^2} E_{\text{samp}}[\mathbf{Q}(\cdot, \cdot), \mathbf{P}(\cdot, \cdot)] \right\}$$

$$A \begin{pmatrix} \mathbf{q} \\ \mathbf{p} \end{pmatrix} + A_0 = \begin{pmatrix} a_{11} & \dots & a_{1d} \\ \vdots & \ddots & \vdots \\ a_{d1} & \dots & a_{dd} \end{pmatrix} \begin{pmatrix} \mathbf{q} \\ \mathbf{p} \end{pmatrix} + \begin{pmatrix} a_{01} \\ \vdots \\ a_{0d} \end{pmatrix}$$

変換  $\mathbf{Q}(\cdot, \cdot), \mathbf{P}(\cdot, \cdot)$  を，線形変換  
 $(\mathbf{Q}, \mathbf{P})^t = A(\mathbf{q}, \mathbf{p})^t$   
 に限定して考える。

$$P(a_{01}, \dots, a_{11}, \dots, a_{2d 2d}) = \frac{1}{Z} \exp \left[ -\frac{N}{2\sigma^2} E_{\text{samp}}(a_{01}, \dots, a_{11}, \dots, a_{2d 2d}) \right]$$

## ●ベイズ推論的枠組みによる変換の探索範囲の制約

$$P(a_{01}, a_{02}, a_{11}, \dots, a_{2d 2d}) = \frac{1}{Z} \exp \left[ -\frac{N}{2\sigma^2} E_{\text{samp}}(a_{01}, a_{02}, a_{11}, \dots, a_{2d 2d}) \right]$$

$$A\mathbf{x} + A_0 = \begin{pmatrix} a_{11} & \dots & a_{1d} \\ \vdots & \ddots & \vdots \\ a_{d1} & \dots & a_{dd} \end{pmatrix} \mathbf{x} + \begin{pmatrix} a_{01} \\ \vdots \\ a_{0d} \end{pmatrix}$$

物理的な制約条件・事前知識

([step2]に対応)

$$P(a_{01}, a_{02}, a_{11}, \dots, a_{2d 2d}) = \frac{1}{Z} \exp \left[ -\frac{N}{2\sigma^2} E_{\text{samp}}(a_{01}, a_{02}, a_{11}, \dots, a_{2d 2d}) \right] \times q(a_{01}, \dots, a_{2d 2d})$$

$$q(a_{01}, \dots, a_{2d 2d}) = \begin{cases} \text{const.} & \text{for } \det A = 1 \\ 0 & \text{for } \det A \neq 1 \end{cases}$$

▶ リウビルの定理 (体積保存則)

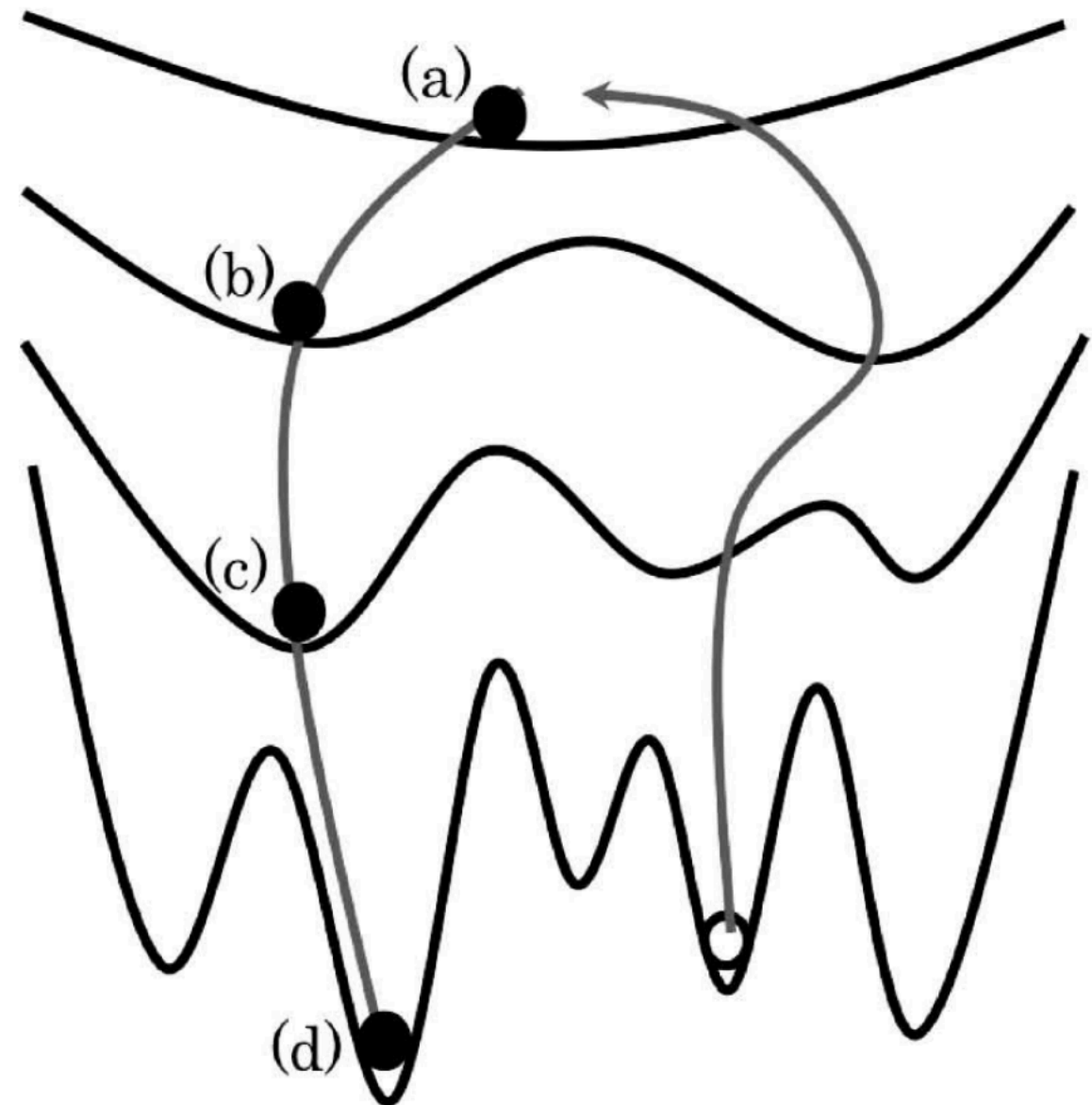
$$P(a_{11}, a_{12}, a_{21}, \dots, a_{2d2d}) = \frac{1}{Z} \exp \left[ -\frac{N}{2\sigma^2} E_{\text{samp}}(a_{11}, a_{12}, a_{21}, \dots, a_{2d2d}) \right]$$

- 交換モンテカルル口法 [Hukushima and Nemoto 1995]

$$P(A^l) = \frac{1}{Z} \exp \left[ -\beta \frac{N}{2\sigma^2} E_{\text{samp}}(A^l) \right]$$

以下の同時分布からMCMCサンプリング

$$P(A^1, A^2 \dots A^L) = \prod_{l=1}^L P(A^l)$$

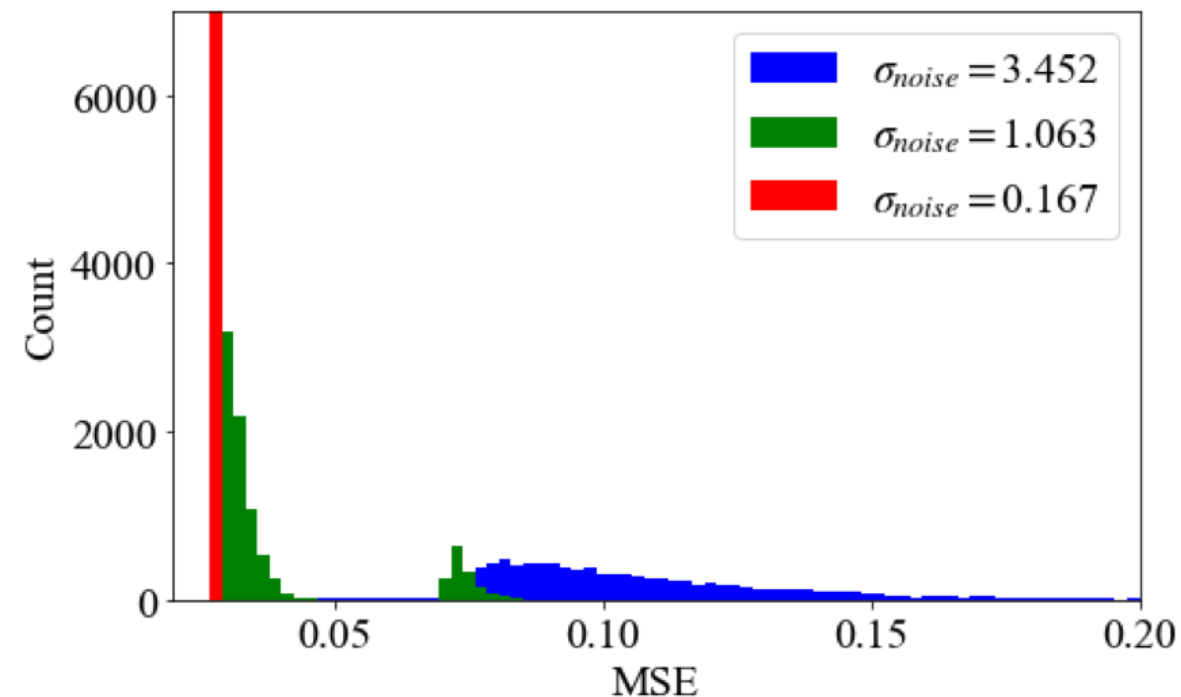


[永田 and 岡田 15]

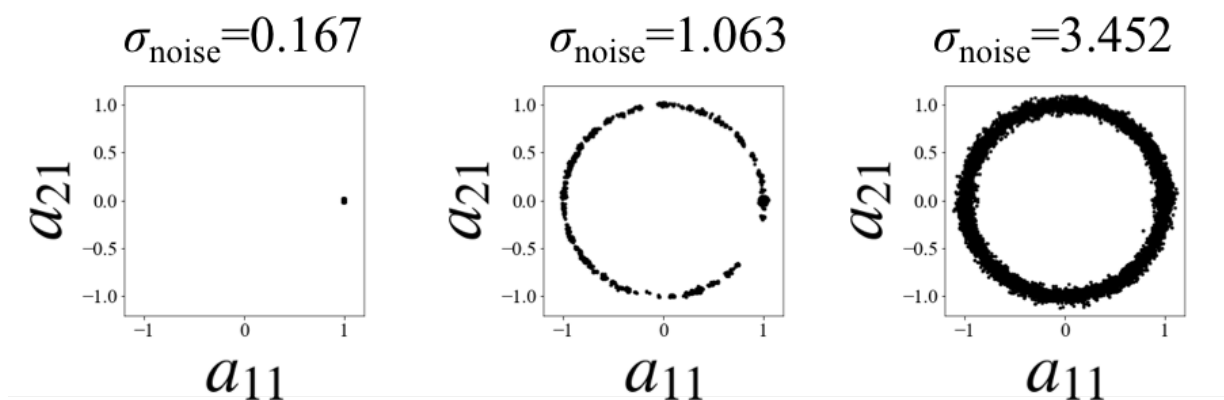
[Y. Mototake, Phys. Rev. E, 103, 033303, 2021]

$$P(a_{11}, a_{12}, a_{21}, \dots, a_{2d 2d}) = \frac{1}{Z} \exp \left[ -\frac{N}{2\sigma^2} E_{\text{samp}}(a_{11}, a_{12}, a_{21}, \dots, a_{2d 2d}) \right]$$

- 交換モンテカルロ法 [Hukushima and Nemoto 1995]



(a)



([step3]に対応)

---

**Algorithm 1** Estimation of the invariant transformation set

---

**Input:** dataset  $D = \{\mathbf{q}_{t_i}^i, \mathbf{p}_{t_i}^i, \mathbf{q}_{t_i+\Delta t}^i, \mathbf{p}_{t_i+\Delta t}^i\}_{i=1}^N$  in a given coordinate system.

**Output:** Invariant transformation set  $D_a = \{(a_{11}, a_{12} \cdots, a_{1d}, a_{21} \cdots, a_{2d} \cdots, a_{2d} \cdots, a_{2d} \cdots)\}_{n_a=1}^{N_a}$ .

**Step 1:** Train the deep autoencoder with dataset  $D$ .

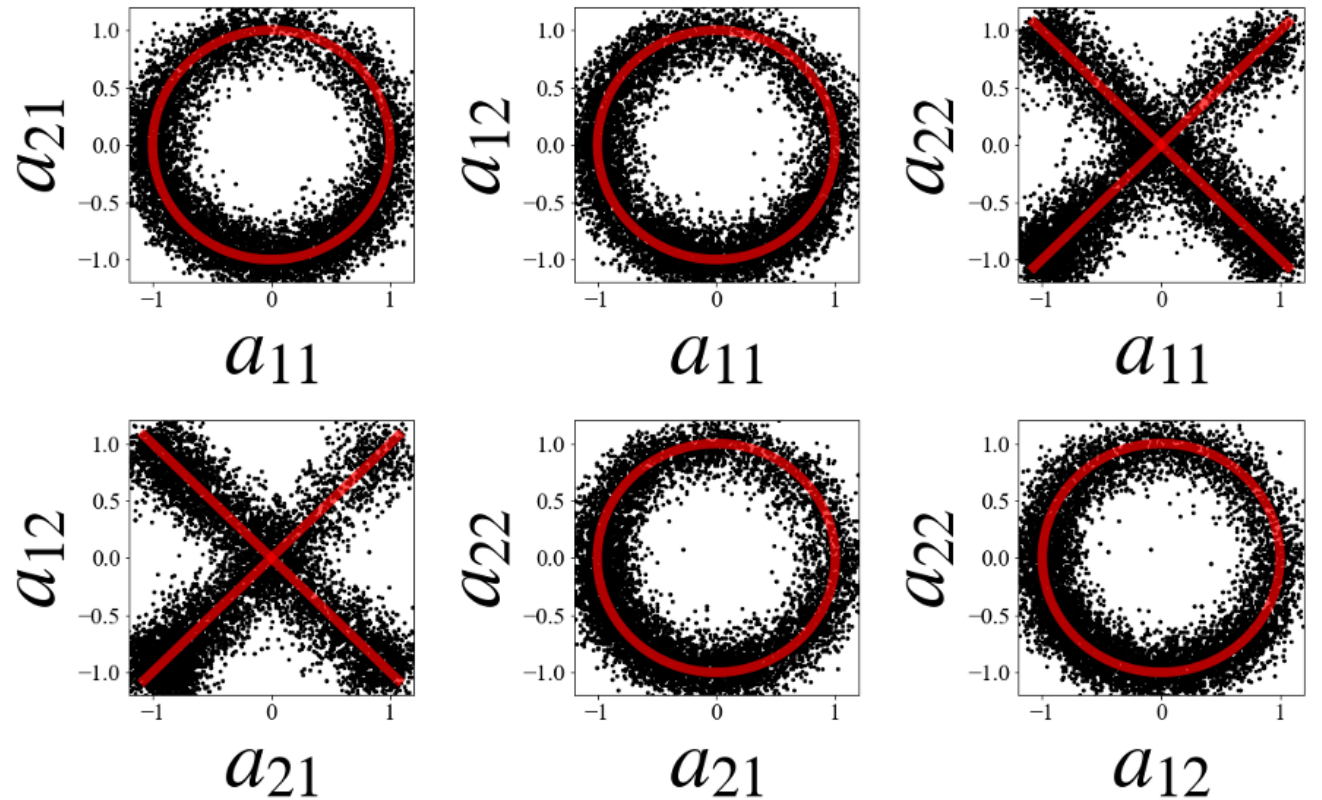
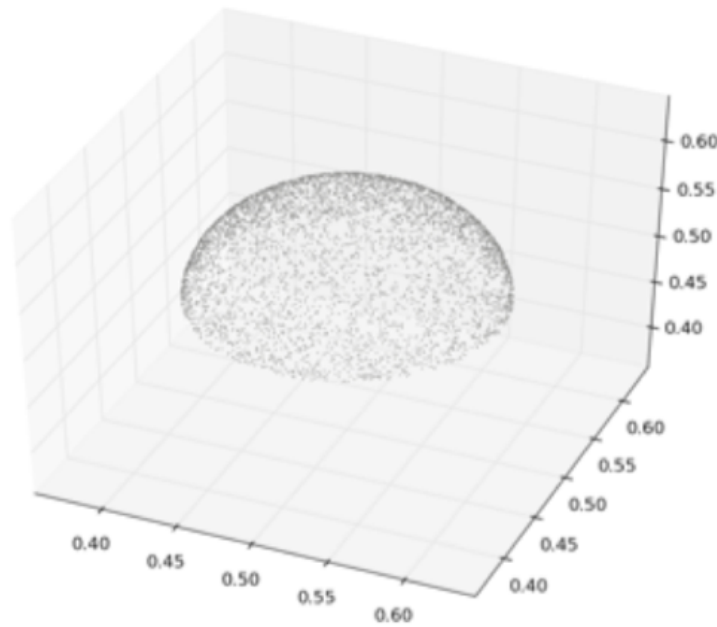
**Step 2:** Using the trained deep autoencoder and REMC method, sampling transformation parameters  $a_{11}, a_{12}, a_{21}, \cdots, a_{2d} \cdots$  from multiple probability distributions  $P'(a_{11}, a_{12}, a_{21}, \cdots, a_{2d} \cdots)$  corresponding to different noise intensities  $\sigma'$ .

**Step 3:** Select  $\sigma'$  from the distribution structure of the sampling results and output the sampling result of the selected  $\sigma'$  state as  $D_a$ .

---

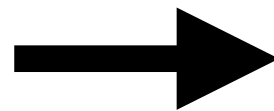


対称な変換の集合  $D_a := \{(a_{01}, a_{02}, \dots, a_{0, 2d}, a_{11}, \dots, a_{2d, 2d})_{n_a}\}_{n_a=1}^{N_a}$



$$(\delta q_{ij}, \delta p_{ij}) = \left( \varepsilon \frac{\partial Q_i(q, p, \theta)}{\partial \theta_j} \Big|_{\theta=\vec{0}}, \varepsilon \frac{\partial P_i(q, p, \theta)}{\partial \theta_j} \Big|_{\theta=\vec{0}} \right)$$

無限小変換

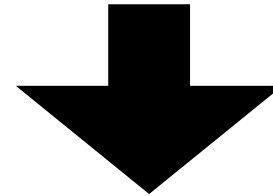


$$M_{\text{invariant}} \sim \left\{ A(\theta) \begin{pmatrix} q \\ p \end{pmatrix} + A_0(\theta) \mid \theta \in \mathbb{R}^{d_\theta} \right\}$$

リー群のなす多様体の接空間  $T_I M_{\text{invariant}}$

## 対称変換の集合

$$D_a := \{(a_{01}, a_{02}, \dots, a_{0, 2d}, a_{11}, \dots, a_{2d, 2d})_{n_a}\}_{n_a=1}^{N_a}$$



$$A'(\boldsymbol{\theta}) = (a'_1(\boldsymbol{\theta}), \dots, a'_{d'}(\boldsymbol{\theta})) := (a_{01}(\boldsymbol{\theta}), \dots, a_{0d}(\boldsymbol{\theta}), a_{11}(\boldsymbol{\theta}), \dots, a_{1d}(\boldsymbol{\theta}), \dots, a_{d1}(\boldsymbol{\theta}), \dots, a_{2d, 2d}(\boldsymbol{\theta}))$$

$$\begin{cases} f_1(a'_1, \dots, a'_{d'}) = 0 \\ \vdots \\ f_{d'-d_\theta}(a'_1, \dots, a'_{d'}) = 0 \end{cases}$$

## リー群のなす多様体の陰関数表示

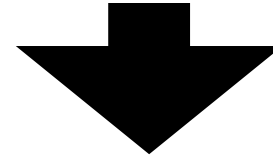


$a_{ij}$  の一部を変換のパラメータ  $\theta$  とする →

陰関数定理より

ただし,  $J_{kl} = \frac{\partial f_k(a'_1, \dots, a'_{d'})}{\partial b_l}$

が正則であるとする.



$$(b_1, b_2, \dots, b_{d_\theta}) \subset A', \quad \{c_k\}_{k=1}^{d'-d_\theta} := A' \setminus \{b_l\}_{l=1}^{d_\theta}, \quad c_k = g_i(b_1, \dots, b_{d_\theta})$$

$$\begin{cases} h_1(c_1, b_1, \dots, b_{d_\theta}) = 0 \\ \vdots \\ h_{d'-d_\theta}(c_{d'-d_\theta}, b_1, \dots, b_{d_\theta}) = 0 \end{cases}$$

$b$ を $\theta$ のかわりの連続パラメータと考えると，無限小変換を推定

$$\left\{ \begin{array}{l} \frac{\partial}{\partial b_l} h_1(c_1, b_1, \dots, b_{d_\theta}) \Big|_{A'=e_I} = 0 \\ \vdots \\ \frac{\partial}{\partial b_l} h_{d'-d_\theta}(c_{d'-d_\theta}, b_1, \dots, b_{d_\theta}) \Big|_{A'=e_I} = 0 \end{array} \right.$$

$$\begin{pmatrix} \delta q_l \\ \delta p_l \end{pmatrix} = \varepsilon \frac{A(b_l)}{\partial b_l} \Big|_{A=I} \begin{pmatrix} \mathbf{q} \\ \mathbf{p} \end{pmatrix} + \varepsilon \frac{A_0(b_l)}{\partial b_l} \Big|_{A_0=\mathbf{0}} = \varepsilon \begin{pmatrix} \frac{\partial a_{11}}{\partial b_l} \Big|_{A=I} & \cdots & \frac{\partial a_{d1}}{\partial b_l} \Big|_{A=I} \\ \vdots & \ddots & \vdots \\ \frac{\partial a_{1d}}{\partial b_l} \Big|_{A=I} & \cdots & \frac{\partial a_{dd}}{\partial b_l} \Big|_{A=I} \end{pmatrix} \begin{pmatrix} \mathbf{q} \\ \mathbf{p} \end{pmatrix} + \varepsilon \begin{pmatrix} \frac{\partial a_{01}}{\partial b_l} \Big|_{A=I} \\ \vdots \\ \frac{\partial a_{0d}}{\partial b_l} \Big|_{A=I} \end{pmatrix}$$

サンプリング結果  $D_b \equiv \{(c_k, b_1, b_2, \dots, b_{d'_\theta})_{n_a}\}_{n_a=1}^{N_a}$  を,

$$\hat{h}_k(c_k, b_1, b_2, \dots, b_{d'_\theta}; \beta, \gamma, d'_\theta) := \sum_{s_0=0}^{d_b} \sum_{s_1=0}^{d_b} \cdots \sum_{s_{d'_\theta}=0}^{d_b} \gamma_{s_0 s_1 s_2 \dots s_{d'_\theta}} \beta_{s_0 s_1 s_2 \dots s_{d'_\theta}} c_k^{s_0} b_1^{s_1} b_2^{s_2} \cdots b_{d'_\theta}^{s_{d'_\theta}} = 0$$

で直交距離回帰して、多様体の陰関数表示を得る.

$$\begin{cases} h_1(c_1, b_1, \dots, b_{d_\theta}) = 0 \\ \vdots \\ h_{d'-d_\theta}(c_{d'-d_\theta}, b_1, \dots, b_{d_\theta}) = 0 \end{cases}$$

ここで、 $\beta$  は回帰係数、 $\gamma$  は基底選択を表現するバイナリベクトルとする.

$\gamma$  は **ベイズ情報量基準 (BIC)** で決定する.

**([step3]に対応)**

---

## Algorithm 2 Estimation of infinitesimal transformation

---

**Input:** Sampling results of Method 1,  $D_a = \{(a_{11}, a_{12} \cdots, a_{1d}, a_{21} \cdots, a_{2d} \cdots, a_{2d} \cdots)\}_{n_a=1}^{N_a}$ , and  $d'_\theta$ .

**Output:** Infinitesimal transformation,  $\delta \mathbf{q}_l, \delta \mathbf{p}_l$ .

**Step 1:** Extract  $D_b = \{(c_k, b_1, b_2, \cdots, b_{d'_\theta})_{n_a}\}_{n_a=1}^{N_a}$  from  $D_a$ .

**Step 2:** Fit  $D_b$  with the implicit polynomial function  $\hat{h}_k(c_k, b_1^l, b_2^l, \cdots, b_{d'_\theta}^l; \beta, \gamma, d'_\theta)$  [Eq. (54)] for each  $c_k$ .

**Step 3:** Estimate the likelihood [Eq. (G1)] by numerical integration of  $Z$  [Eq. (G2)].

**Step 4:** Select the indicator vector  $\gamma$  and the dimension  $d'_\theta$  of  $M_{\text{invariant}}$  in Eq. (54) for each  $c_k$  using the BIC,

**Step 5:** Determine whether the Jacobi matrix  $J_{kl} = \frac{\partial h_k(c_k, b_1, \cdots, b_{d'_\theta})}{\partial b_l}$  is nonsingular. If  $J_{kl}$  is singular, return to Step 1 and re-extract  $D'_b$ .

**Step 6:** Differentiate the obtained simultaneous equations with respect to  $b_l$  around a point  $e_I$  to obtain Eq. (52).

**Step 7:** Solve the simultaneous equations in Eq. (52) and obtain the infinitesimal transformation,  $\delta \mathbf{q}_l, \delta \mathbf{p}_l$ .

---



## Definitions

$$H: M_E \rightarrow \mathbb{R}$$

$$H': M'_E \rightarrow \mathbb{R}$$

$$M_E := \{\mathbf{q}_t, \mathbf{p}_t \mid H(\mathbf{q}_t, \mathbf{p}_t) = E\}$$

$$M'_E := \{\mathbf{Q}_T, \mathbf{P}_T \mid H'(\mathbf{Q}_T, \mathbf{P}_T) = E\}$$

$$\mathbb{U}: (\mathbf{q}_t, \mathbf{p}_t) \mapsto (\mathbf{q}_{t+1}, \mathbf{p}_{t+1})$$

$$\mathbb{U}': (\mathbf{Q}_t, \mathbf{P}_t) \mapsto (\mathbf{Q}_{t+1}, \mathbf{P}_{t+1})$$

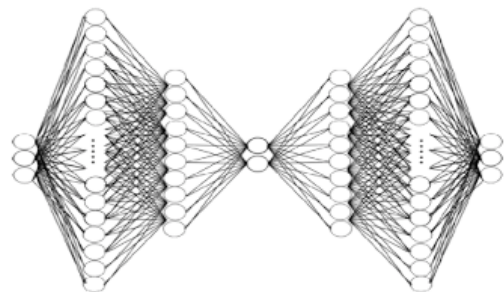
$$\mathbb{C}: (\mathbf{q}_t, \mathbf{p}_t, \mathbf{q}_{t+1}, \mathbf{p}_{t+1}) \mapsto (\mathbf{Q}_T, \mathbf{P}_T, \mathbf{Q}_{T+1}, \mathbf{P}_{T+1})$$

$$f_{\text{DNN}}(x) = \begin{cases} x & \text{if } x \in M_E \times \mathbb{U}(M_E) \\ y & \text{else} \end{cases}$$

where  $y \neq x$ .

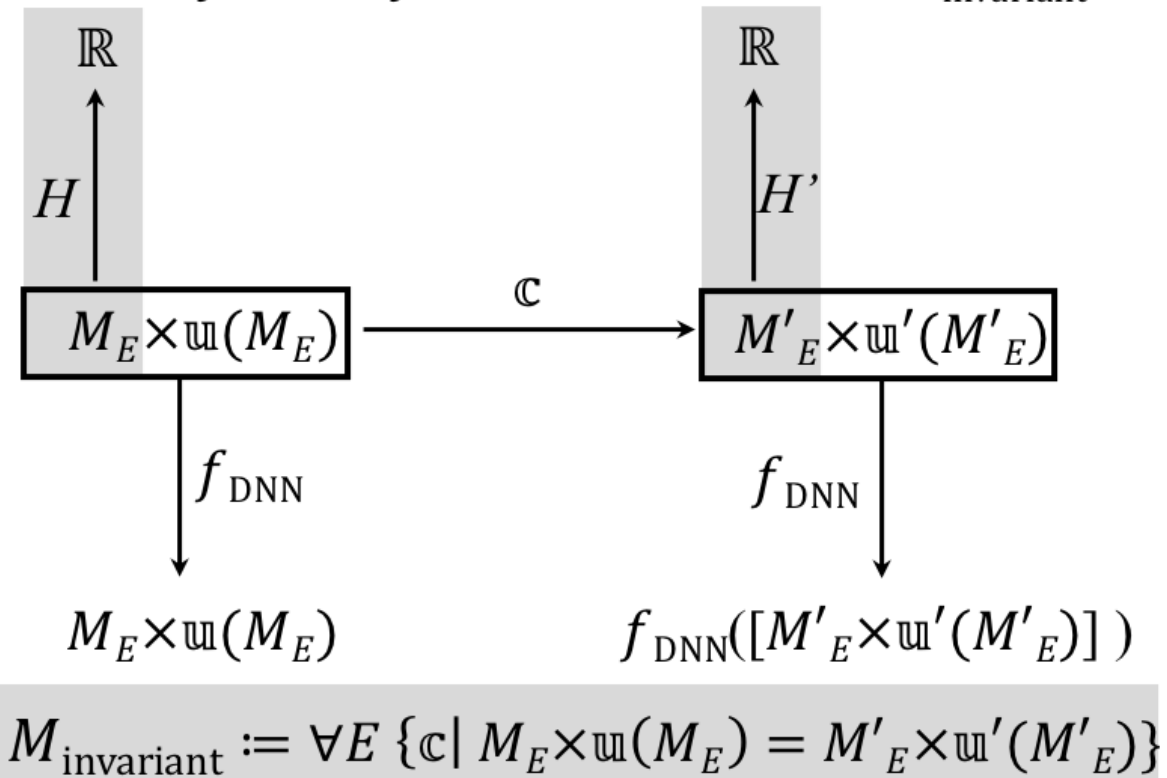
$T_l M$ : tangent space of  $M$  at identity mat.

$G_l$ :  $\mathbb{R}$  (conserved value)



$f_{\text{DNN}}$ : deep auto encoder trained by  $M_E \times \mathbb{U}(M_E)$

Relationship between time series dataset  $M_E \times \mathbb{U}(M_E)$  and symmetry transformation set  $M_{\text{invariant}}$



【Method 1: Inferring the symmetry (Sec. III A)】

$$M_{\text{invariant}} \approx \forall E \{C \mid [M'_E \times \mathbb{U}'(M'_E)] = f_{\text{DNN}}([M'_E \times \mathbb{U}'(M'_E)])\}$$

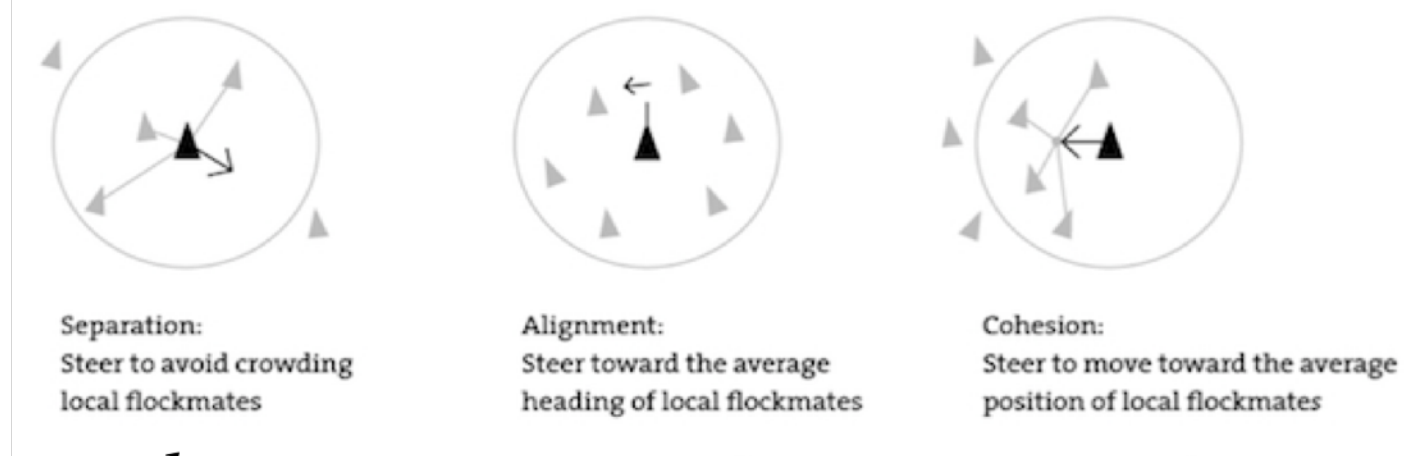


【Method 2: Inferring the conservation law (Sec. III B)】

$$[T_l M_{\text{invariant}}]_l = \left( \frac{\partial G_l}{\partial p}, -\frac{\partial G_l}{\partial q} \right) \quad \text{Derived from Noether's theorem. (Sec. II)}$$

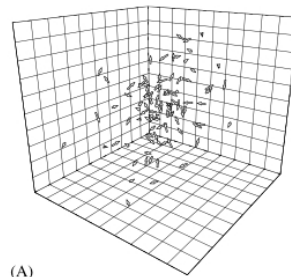
## モデル

Reynolds model [Reynolds et al. 86]

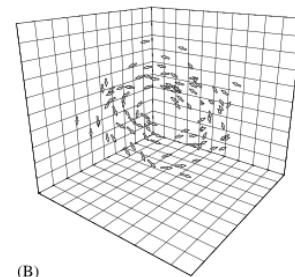


$$\frac{dv}{dt} = \vec{F}_{separation} + \vec{F}_{alignment} + \vec{F}_{cohesion}$$

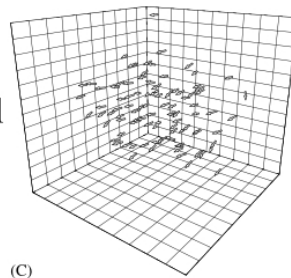
Swarm



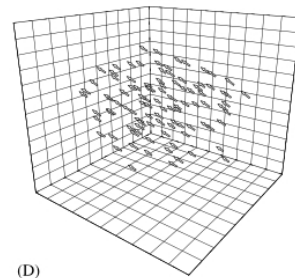
Torus



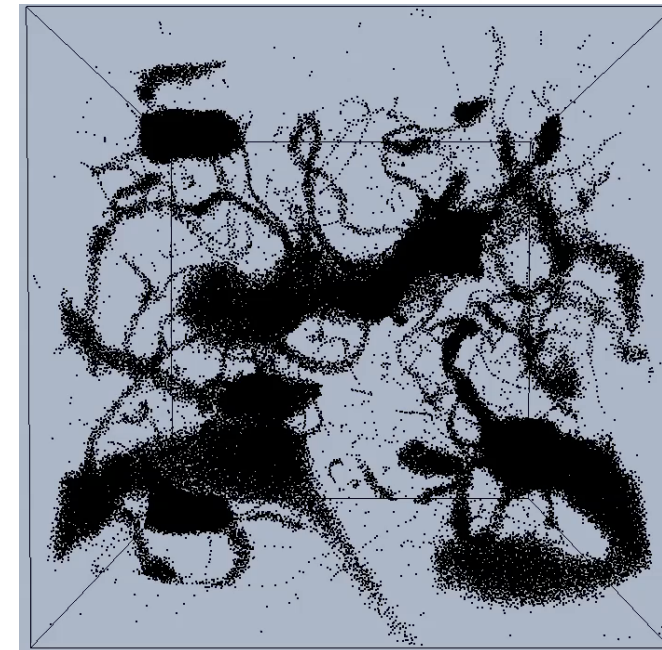
Dynamic Parallel



Highly Parallel



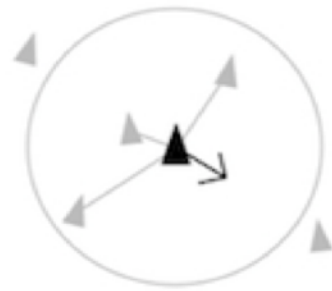
[Couzin et.al 2002]



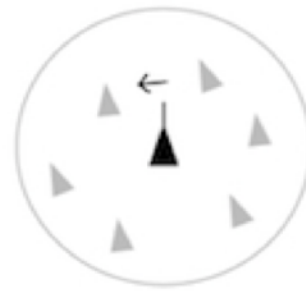
[Mototake et.al 2015]

## モデル

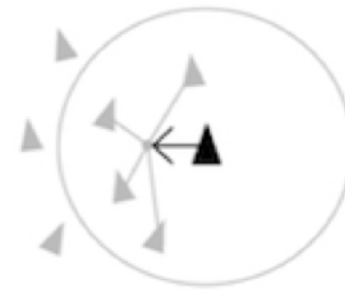
Reynolds model [Reynolds et al. 86]



Separation:  
Steer to avoid crowding  
local flockmates



Alignment:  
Steer toward the average  
heading of local flockmates



Cohesion:  
Steer to move toward the average  
position of local flockmates

$$\frac{dv}{dt} = \vec{F}_{separation} + \vec{F}_{alignment} + \vec{F}_{cohesion}$$



## 座標系

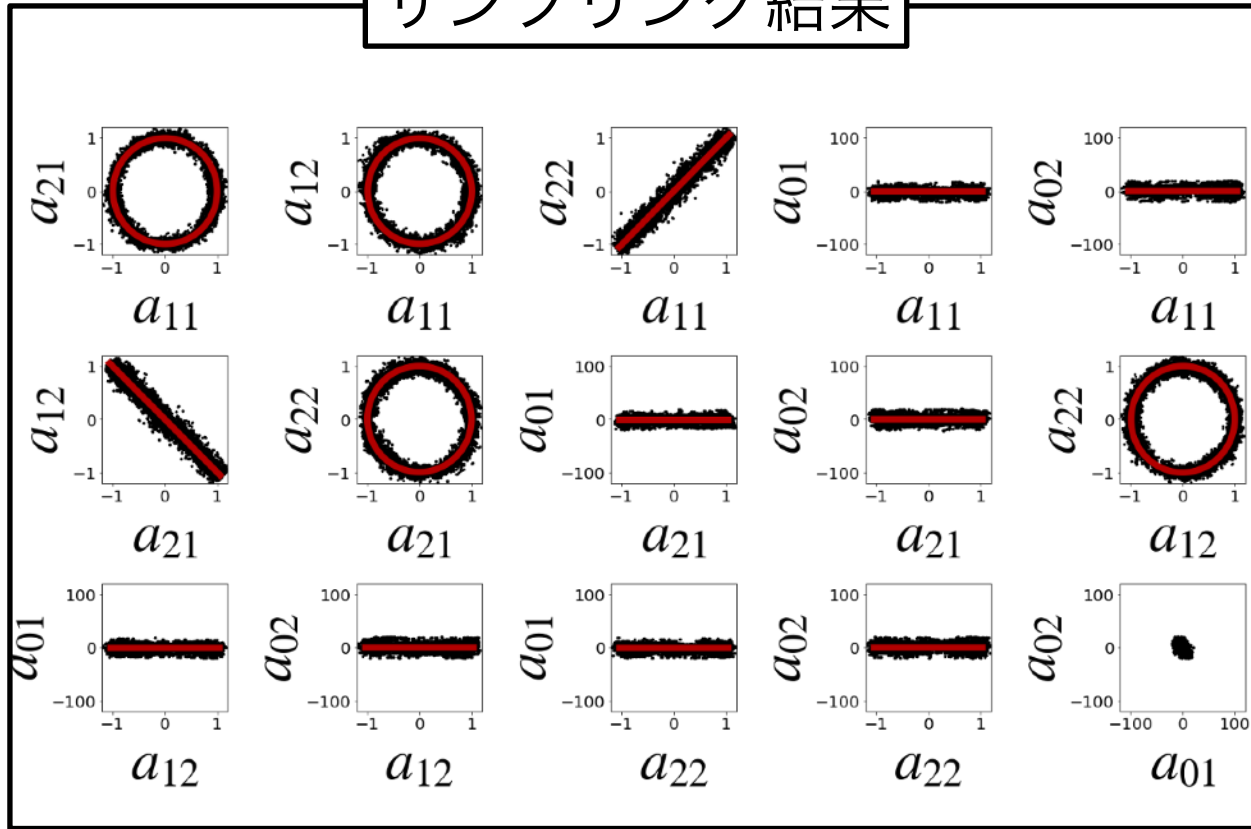
$$D = \{ \tilde{\mathbf{q}}(t_i)_i, \tilde{\mathbf{p}}(t_i)_i, \tilde{\mathbf{q}}(t_i + \Delta t)_i, \tilde{\mathbf{p}}(t_i + \Delta t)_i \}_{i=1}^{N_R T}$$

$$(\tilde{\mathbf{q}}, \tilde{\mathbf{p}}) = (q_1 - \bar{q}_1, q_2 - \bar{q}_2, p_1 - \bar{p}_1, p_2 - \bar{p}_2)$$

## 変換行列

$$\begin{pmatrix} a_{11} & 1.0 a_{21} & 0 & 0 \\ 1.0 a_{12} & a_{22} & 0 & 0 \\ 0 & 0 & a_{22} & -1.0 a_{12} \\ 0 & 0 & -1.0 a_{21} & a_{11} \end{pmatrix} \begin{pmatrix} q_1 \\ q_2 \\ p_1 \\ p_2 \end{pmatrix} + \begin{pmatrix} a_{01} \\ a_{02} \\ 0.167 a_{02} \\ -0.161 a_{01} \end{pmatrix}$$

## サンプリング結果



## 陰関数の回帰結果

$$\begin{cases} 1.01 a_{11}^2 + 1.01 a_{21}^2 - 0.01 a_{11} + 0.00 a_{21} + 0.00 a_{11}a_{21} = 1.00 \\ 1.01 a_{11}^2 + 1.00 a_{12}^2 - 0.01 a_{11} = 0.99 \\ 1.00 a_{11} - 1.00 a_{22} + 0.00 a_{22} = 0 \\ 1.00 a_{21} - 1.00 a_{12} = 0 \\ 1.00 a_{21}^2 + 1.00 a_{22}^2 - 0.01 a_{22} = 0.99 \\ 1.02 a_{12}^2 + 1.00 a_{22}^2 - 0.01 a_{12}a_{22} = 1.00 \end{cases}$$

## 無限小変換の推定結果

$$\delta \mathbf{q} = \varepsilon \begin{pmatrix} \frac{\partial a_{11}}{\partial a_{21}} & \frac{\partial a_{21}}{\partial a_{21}} \\ \frac{\partial a_{12}}{\partial a_{21}} & \frac{\partial a_{21}}{\partial a_{21}} \end{pmatrix} \mathbf{q} + \varepsilon \begin{pmatrix} \frac{\partial a_{01}}{\partial a_{21}} \\ \frac{\partial a_{02}}{\partial a_{21}} \end{pmatrix} = \varepsilon \begin{pmatrix} \frac{-1.01 \times 2a_{21}}{1.01 \times 2a_{11} - 0.01} \Big|_{A'=e_{\mathbf{I}}} & 1 \\ \frac{-1.00}{1.00} \Big|_{A'=e_{\mathbf{I}}} & \frac{-1.00 \times 2a_{21}}{1.00 \times 2a_{22} + 0.01} \Big|_{A'=e_{\mathbf{I}}} \end{pmatrix} \mathbf{q} = \begin{pmatrix} 0 & \varepsilon \\ -\varepsilon & 0 \end{pmatrix} \mathbf{q}$$

$$\delta \mathbf{p} = \varepsilon \begin{pmatrix} \frac{\partial a_{11}}{\partial a_{21}} & \frac{\partial a_{21}}{\partial a_{21}} \\ \frac{\partial a_{12}}{\partial a_{21}} & \frac{\partial a_{21}}{\partial a_{21}} \end{pmatrix} \mathbf{p} + \varepsilon \begin{pmatrix} \frac{\partial a_{01}}{\partial a_{21}} \\ \frac{\partial a_{02}}{\partial a_{21}} \end{pmatrix} = \begin{pmatrix} 0 & \varepsilon \\ -\varepsilon & 0 \end{pmatrix} \mathbf{p}$$

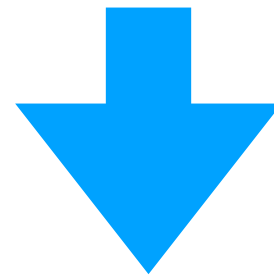
先行研究 [Couzin et.al 2002]  
で示された結果と一致➡

### 保存則

$$G_{\delta} = \varepsilon(x_1 p_2 - x_2 p_1)$$

- 原理的には**非線形変換にも適用可能**.

$$\begin{pmatrix} \delta q_l \\ \delta p_l \end{pmatrix} = \varepsilon \frac{A(b_l)}{\partial b_l} \Big|_{A=I} \begin{pmatrix} \mathbf{q} \\ \mathbf{p} \end{pmatrix} + \varepsilon \frac{A_0(b_l)}{\partial b_l} \Big|_{A_0=0} = \varepsilon \begin{pmatrix} \frac{\partial a_{11}}{\partial b_l} \Big|_{A=I} & \cdots & \frac{\partial a_{d1}}{\partial b_l} \Big|_{A=I} \\ \vdots & \ddots & \vdots \\ \frac{\partial a_{1d}}{\partial b_l} \Big|_{A=I} & \cdots & \frac{\partial a_{dd}}{\partial b_l} \Big|_{A=I} \end{pmatrix} \begin{pmatrix} \mathbf{q} \\ \mathbf{p} \end{pmatrix} + \varepsilon \begin{pmatrix} \frac{\partial a_{01}}{\partial b_l} \Big|_{A=I} \\ \vdots \\ \frac{\partial a_{0d}}{\partial b_l} \Big|_{A=I} \end{pmatrix}$$



$$(\delta \mathbf{q}_l, \delta \mathbf{p}_l) = \varepsilon \left( \frac{\partial Q(\mathbf{q}, \mathbf{p}; \mathbf{a}(\theta_l))}{\partial \theta_l} \Big|_{\theta_l=0}, \frac{\partial P(\mathbf{q}, \mathbf{p}; \mathbf{a}(\theta_l))}{\partial \theta_l} \Big|_{\theta_l=0} \right)$$

$$(\delta \mathbf{q}_l, \delta \mathbf{p}_l) = \varepsilon \left( \sum_{k=1}^{d_a} \frac{\partial Q(\mathbf{q}, \mathbf{p}; \mathbf{a})}{\partial a_k} \frac{\partial a_k(\boldsymbol{\theta})}{\partial \theta_l} \Big|_{\boldsymbol{\theta}=0}, \sum_{k=1}^{d_a} \frac{\partial P(\mathbf{q}, \mathbf{p}; \mathbf{a})}{\partial a_k} \frac{\partial a_k(\boldsymbol{\theta})}{\partial \theta_l} \Big|_{\boldsymbol{\theta}=0} \right)$$

**▶ いかに変換関数Q,Pを適切に設定できるかが重要**



未知の保存量・対称性, あるいは推定が困難とされる保存量・対称性の推定を実現.

ルンゲ・レンツベクトル

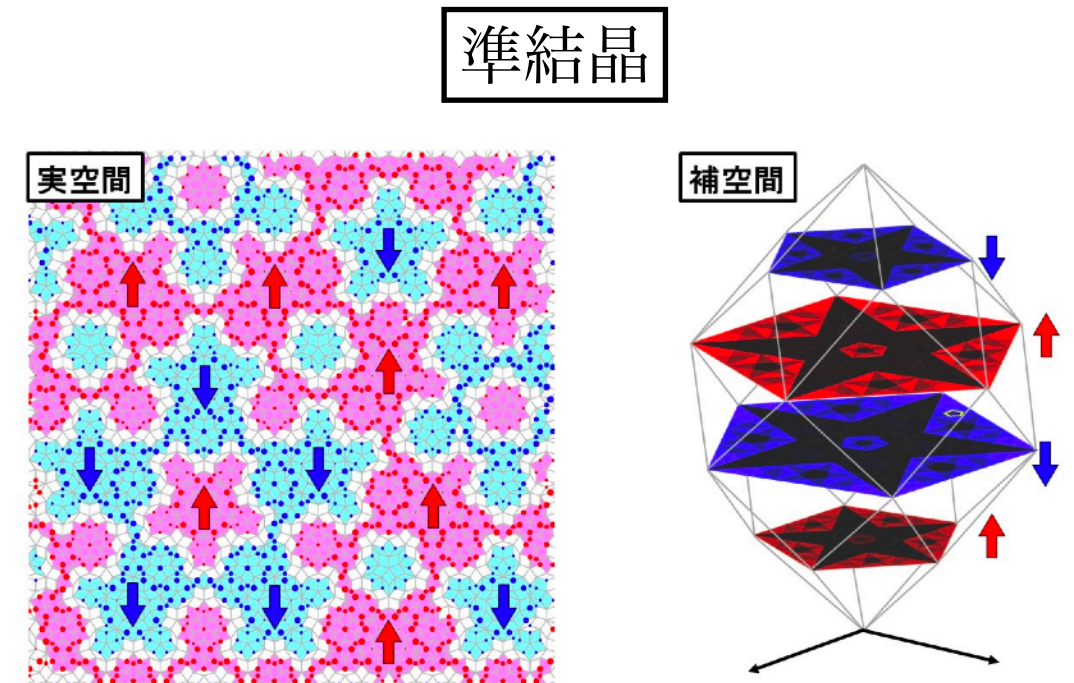
$$\vec{A} = \mathbf{p} \times \mathbf{L} - mG \frac{\mathbf{q}}{\|\mathbf{q}\|_2},$$

$$\mathbf{L} = \mathbf{q} \times \mathbf{p},$$

$$(q_1, q_2, q_3, p_1, p_2, p_3) \rightarrow (\tilde{q}_1, \tilde{q}_2, \tilde{q}_3, \tilde{q}_4, \tilde{p}_1, \tilde{p}_2, \tilde{p}_3, \tilde{p}_4)$$

$$\tilde{\mathbf{q}} = \tilde{\mathbf{q}}(\mathbf{q}, \mathbf{p}) := \frac{\mathbf{q}}{\|\mathbf{q}\|_2} - \frac{\mathbf{q} \cdot \mathbf{p}}{mG} \mathbf{p}, \quad \tilde{q}_4 = \tilde{q}_4(\mathbf{q}, \mathbf{p}) := \frac{p_0}{mG} \mathbf{q} \cdot \mathbf{p},$$

$$\tilde{\mathbf{p}} = \tilde{\mathbf{p}}(\mathbf{q}, \mathbf{p}) := \frac{2p_0 \mathbf{p}}{p_0^2 + p^2}, \quad \tilde{p}_4 = \tilde{p}_4(\mathbf{q}, \mathbf{p}) := \frac{p^2 - p_0^2}{p_0^2 + p^2},$$



[A. Koga and H. Tsunetsugu, Phys. Rev. B 96, (2017) 214402.]

▶ 高次元に射影した空間上で対称性を持つ一群の系がある。

逆二乗則に従う中心力系：

$$H = \frac{\mathbf{p}^2}{2m} + \frac{k}{|\mathbf{r}|}, k : \text{const.}$$

## ルンゲレンツベクトル

$$\mathbf{A} = \mathbf{p} \times \mathbf{L} - mk \frac{\mathbf{q}}{|\mathbf{q}|}$$

$$\mathbf{L} = \mathbf{q} \times \mathbf{p}$$

## SO(4)のリー代数

$$\{L_i, L_j\} = \sum_{k=1}^3 \epsilon_{ijk} L_k$$

$$\{D_i, L_j\} = \sum_{k=1}^3 \epsilon_{ijk} D_k$$

$$\{D_i, D_j\} = \sum_{k=1}^3 \epsilon_{ijk} L_k$$

$$\mathbf{D} = \frac{\mathbf{A}}{\sqrt{-2mH}}$$

## ルンゲレンツベクトルの幾何学的理解

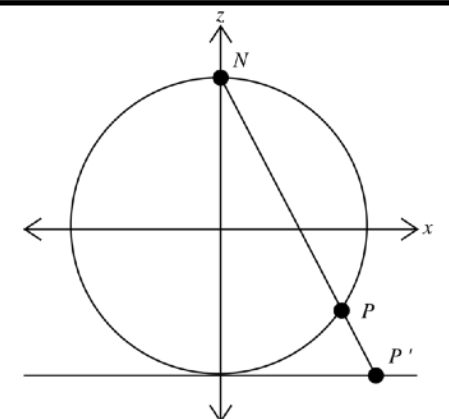
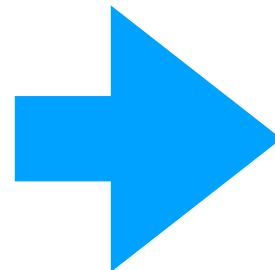
以下のような8次元の位相空間への非線形写像によって、幾何学的により明瞭な空間での理論を構築する

$$(q_1, q_2, q_3, p_1, p_2, p_3) \rightarrow (\tilde{q}_1, \tilde{q}_2, \tilde{q}_3, \tilde{q}_4, \tilde{p}_1, \tilde{p}_2, \tilde{p}_3, \tilde{p}_4)$$

$$\tilde{\mathbf{q}} = \tilde{\mathbf{q}}(\mathbf{q}, \mathbf{p}) := \frac{\mathbf{q}}{\|\mathbf{q}\|_2} - \frac{\mathbf{q} \cdot \mathbf{p}}{mG} \mathbf{p}, \quad \tilde{q}_4 = \tilde{q}_4(\mathbf{q}, \mathbf{p}) := \frac{p_0}{mG} \mathbf{q} \cdot \mathbf{p},$$

$$\tilde{\mathbf{p}} = \tilde{\mathbf{p}}(\mathbf{q}, \mathbf{p}) := \frac{2p_0 \mathbf{p}}{p_0^2 + p^2}, \quad \tilde{p}_4 = \tilde{p}_4(\mathbf{q}, \mathbf{p}) := \frac{p^2 - p_0^2}{p_0^2 + p^2},$$

[H. H. Rogers, J. Math. Phys. 14, 1125 (1973)]

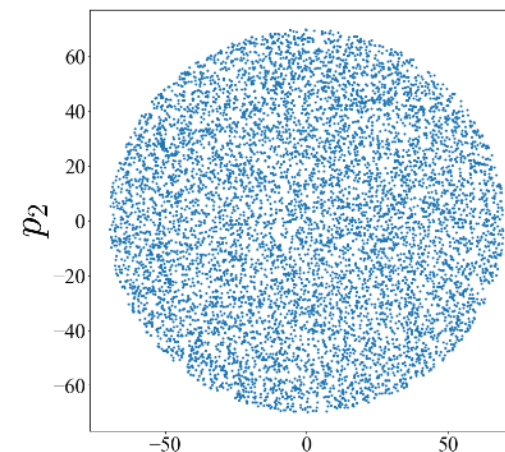
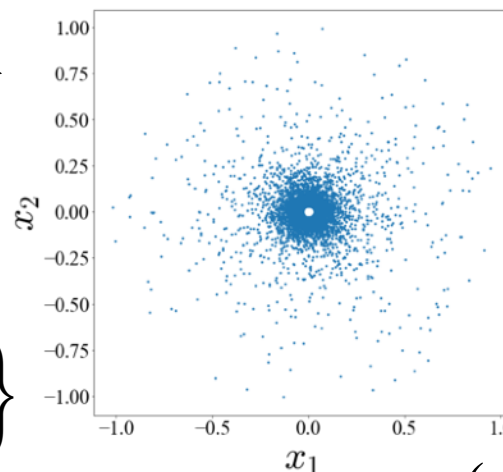


ルンゲレンツベクトルとデータ多様体：

$$H = \frac{\mathbf{p}^2}{2m} + \frac{k}{|\mathbf{r}|}, k : \text{const.}$$

$$\mathbf{q} = (q_1, q_2), \mathbf{p} = (p_1, p_2)$$

$$S_i := \left\{ \mathbf{q}_{t+\Delta t}, \mathbf{p}_{t+\Delta t}, \mathbf{q}_t, \mathbf{p}_t \mid H(\mathbf{q}_t, \mathbf{p}_t) = E_i, \mathbf{p}_{t+\Delta t} = \mathbf{p}_t - \frac{\partial H(\mathbf{q}_t, \mathbf{p}_t)}{\partial \mathbf{q}_t}, \mathbf{q}_{t+\Delta t} = \mathbf{q}_t + \frac{\partial H(\mathbf{q}_t, \mathbf{p}_t)}{\partial \mathbf{p}_t} \right\}$$



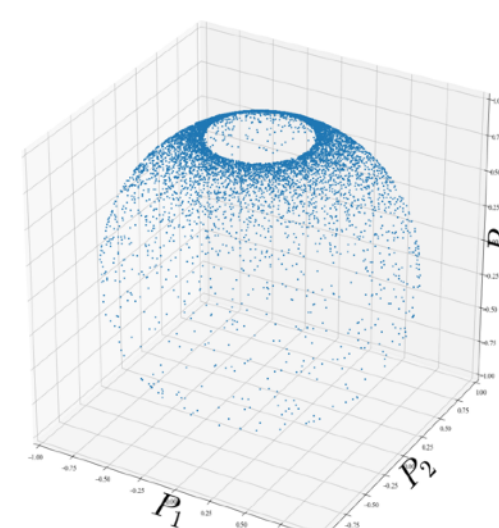
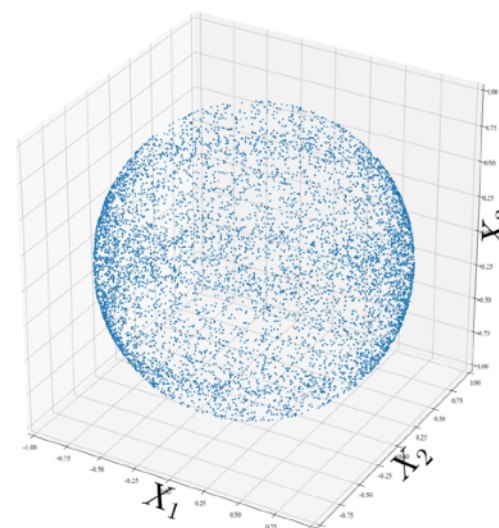
$(q_1, q_2, p_1, p_2)$

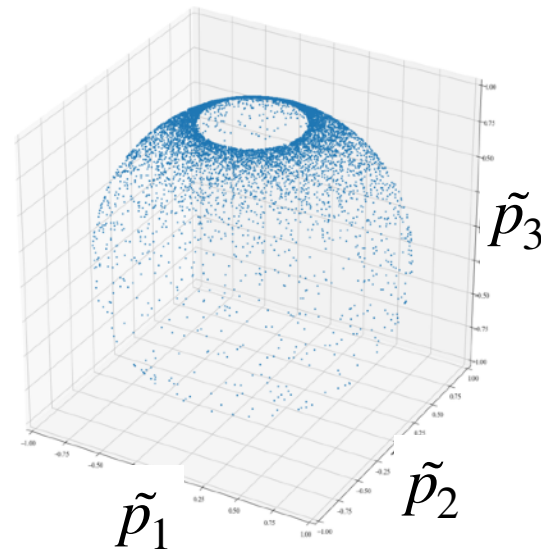
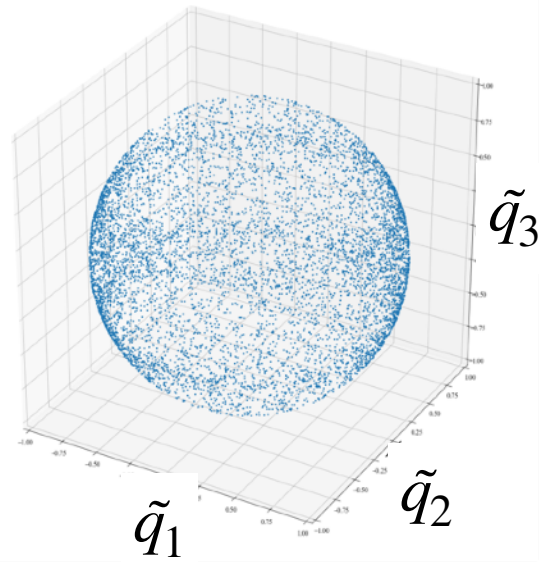
$(q_1, q_2, q_3, p_1, p_2, p_3) \rightarrow (\tilde{q}_1, \tilde{q}_2, \tilde{q}_3, \tilde{q}_4, \tilde{p}_1, \tilde{p}_2, \tilde{p}_3, \tilde{p}_4)$   
 $\tilde{\mathbf{q}} = \tilde{\mathbf{q}}(\mathbf{q}, \mathbf{p}) := \frac{\mathbf{q}}{\|\mathbf{q}\|_2} - \frac{\mathbf{q} \cdot \mathbf{p}}{mG} \mathbf{p}, \quad \tilde{q}_4 = \tilde{q}_4(\mathbf{q}, \mathbf{p}) := \frac{p_0}{mG} \mathbf{q} \cdot \mathbf{p},$   
 $\tilde{\mathbf{p}} = \tilde{\mathbf{p}}(\mathbf{q}, \mathbf{p}) := \frac{2p_0 \mathbf{p}}{p_0^2 + p^2}, \quad \tilde{p}_4 = \tilde{p}_4(\mathbf{q}, \mathbf{p}) := \frac{p^2 - p_0^2}{p_0^2 + p^2},$

[H. H. Rogers, J. Math. Phys. 14, 1125 (1973)]

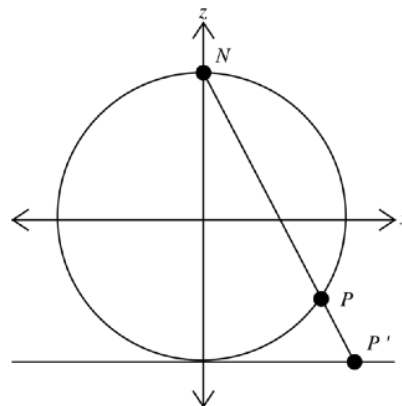
変換

$(\tilde{q}_1, \tilde{q}_2, \tilde{q}_3, \tilde{p}_1, \tilde{p}_2, \tilde{p}_3)$

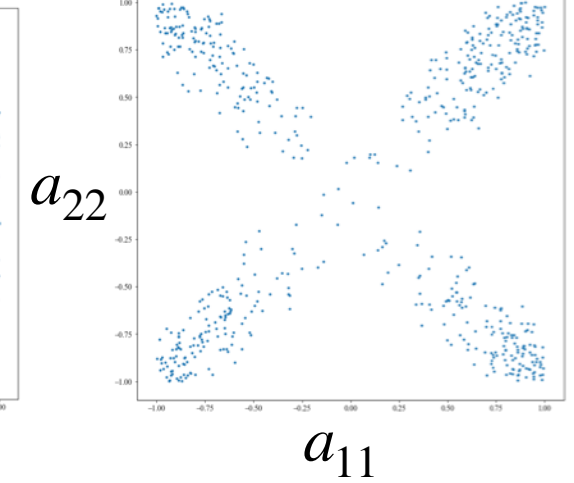
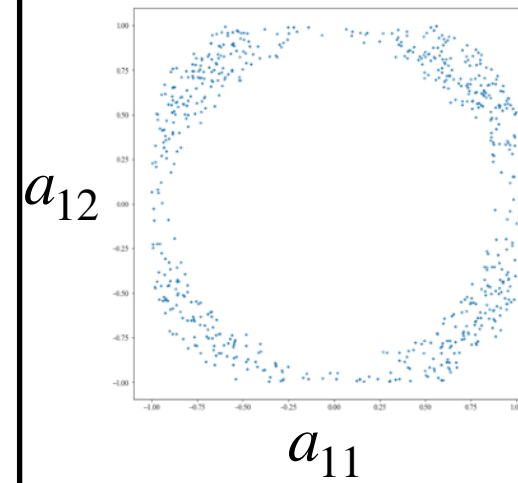




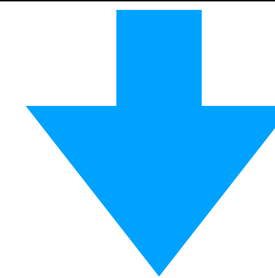
$$\begin{pmatrix} \tilde{Q}_1 \\ \tilde{Q}_2 \\ \tilde{Q}_3 \\ \tilde{P}_1 \\ \tilde{P}_2 \\ \tilde{P}_3 \end{pmatrix} = \begin{pmatrix} a_{11}, a_{12}, 0, 0, 0, 0 \\ a_{21}, a_{22}, 0, 0, 0, 0 \\ 0, 0, 1, 0, 0, 0 \\ 0, 0, 0, a_{11}, a_{12}, 0 \\ 0, 0, 0, a_{21}, a_{22}, 0 \\ 0, 0, 0, 0, 0, 1 \end{pmatrix} \begin{pmatrix} \tilde{q}_1 \\ \tilde{q}_2 \\ \tilde{q}_3 \\ \tilde{p}_1 \\ \tilde{p}_2 \\ \tilde{p}_3 \end{pmatrix}$$



抽出された対称性



► It was suggested that we are possible to extract the SO(4) symmetry corresponding to the Runge-Lenz vector.



**非線形変換をうまく限定できれば  
ルンゲレンツベクトルの推定は実現可能**

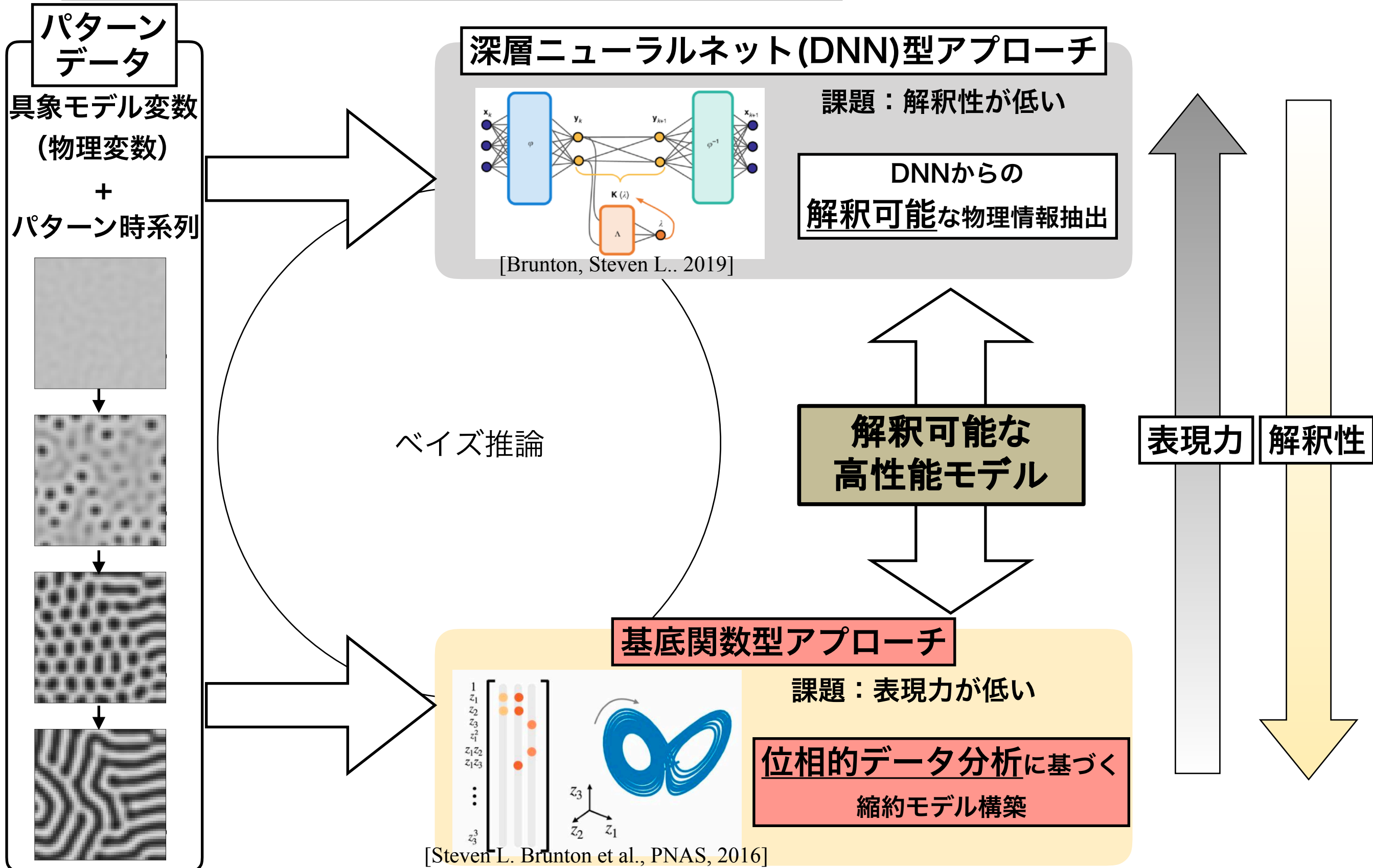


- 時系列データ多様体から、ハミルトン系の対称性の候補が推定可能であることを示した。
- 学習済みのDNNから、時系列データ多様体の対称性を推定できる可能性を示唆した。
- 保存則の数が増えると、ヤコビ行列が正則でなくなる可能性が高まる等、推定が困難になると予想される。
- 位相空間の次元が増えると、サンプリング法による対称性抽出に困難が生じると思われる。
- 力学系時系列データを学習したDNNから、非線形変換に対応する対称性を原理的には抽出可能である



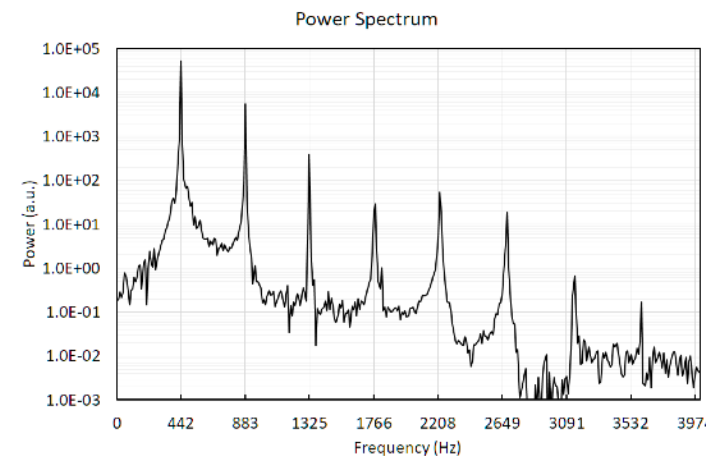
1. 自己紹介
2. 解釈可能AIとデータ駆動理学
3. DNN型アプローチによるデータ駆動理学の実現
- 4. 基底関数型アプローチによるデータ駆動理学の実現**
5. まとめと今後の展望

## 機械学習による解釈可能AI構築アプローチ



**[要件1]** 人間の目で基底空間の特徴構造が検出できる  
(≒可視化として有効である)

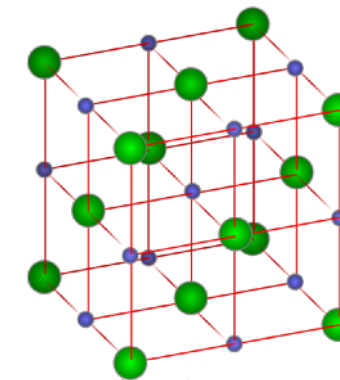
例：フーリエスペクトル



[<https://www.isc.meiji.ac.jp/~be00085/2022ji2/dft/index2.html>]

**[要件2]** 元の空間での構造との対応関係がわかる

例：フーリエスペクトルと周期構造



NaCl結晶(Wikipedia)

**[要件3]** 特徴量としてなんらかの数学的保証がある

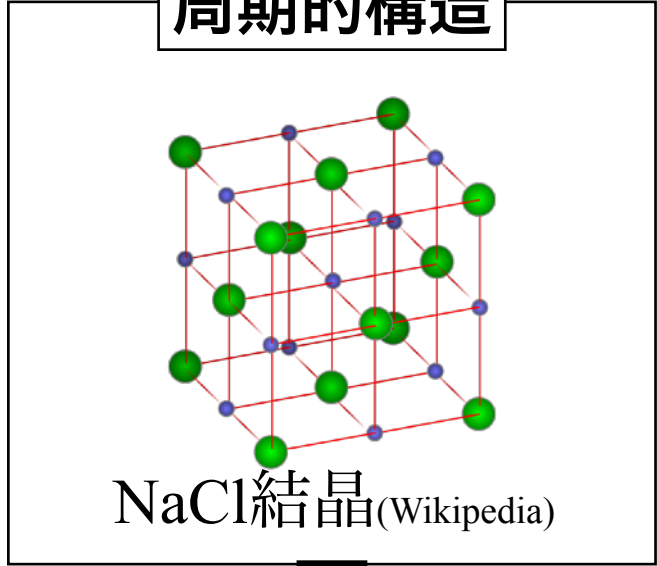
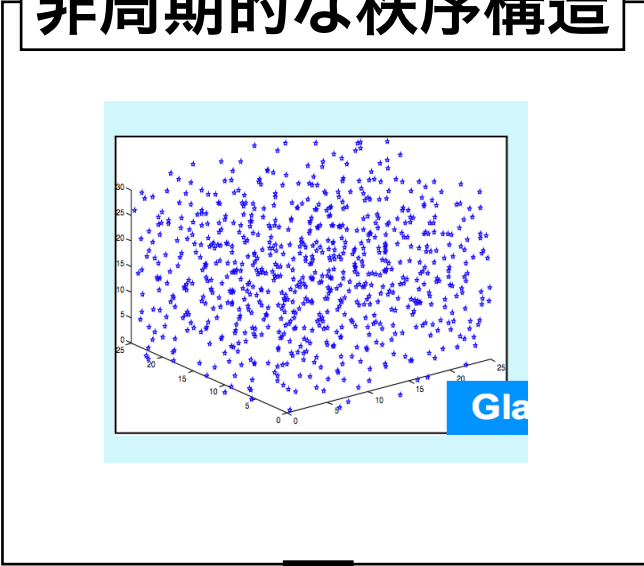
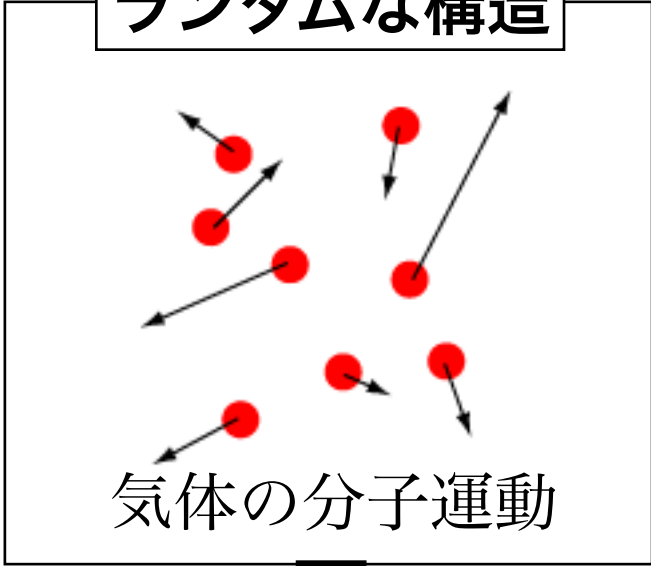
例：フーリエ逆変換による復元保証



ランダムな構造

非周期的な秩序構造

周期的構造

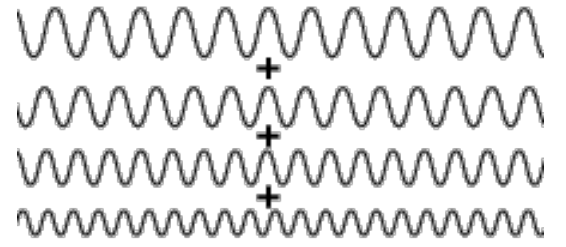
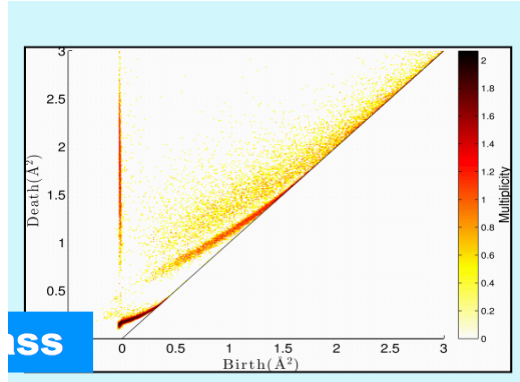


統計量

位相的特徴量

フーリエ基底

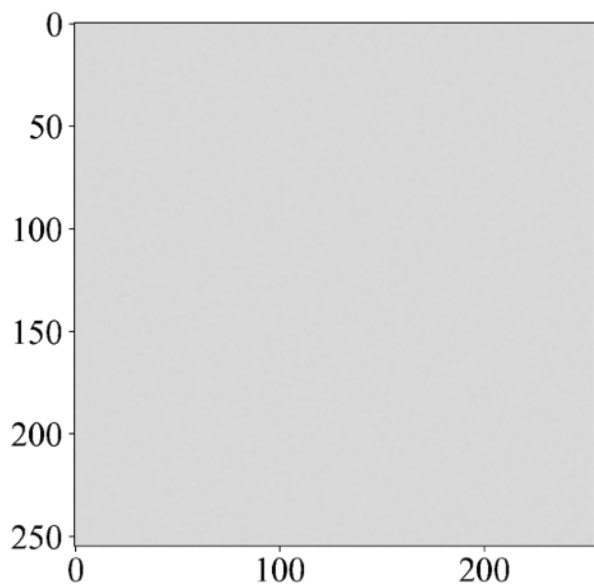
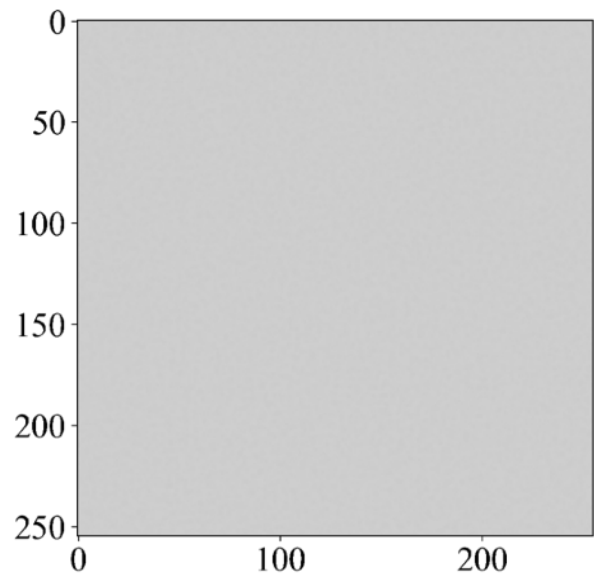
$-S$	U	$V$
H		F
$-p$	G	$T$



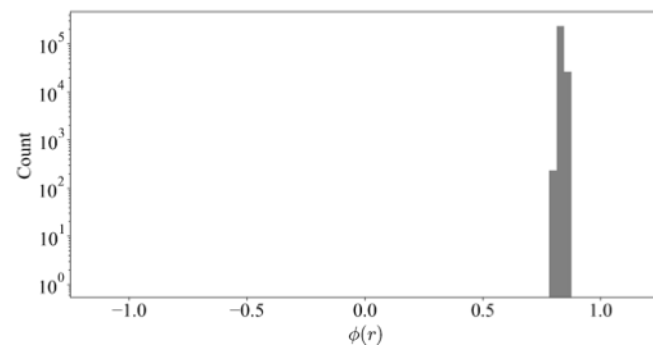
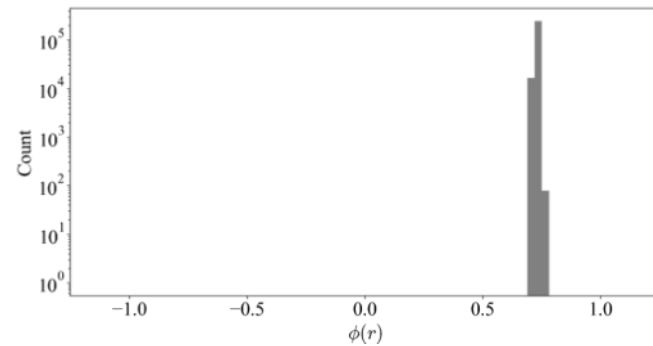
平岡先生スライド: 「パーシステントホモロジーとその応用」  
[http://www.math.chuo-u.ac.jp/ENCwMATH/EwM70\\_Hiraoka.pdf](http://www.math.chuo-u.ac.jp/ENCwMATH/EwM70_Hiraoka.pdf)

<http://k-hiura.cocolog-nifty.com/>

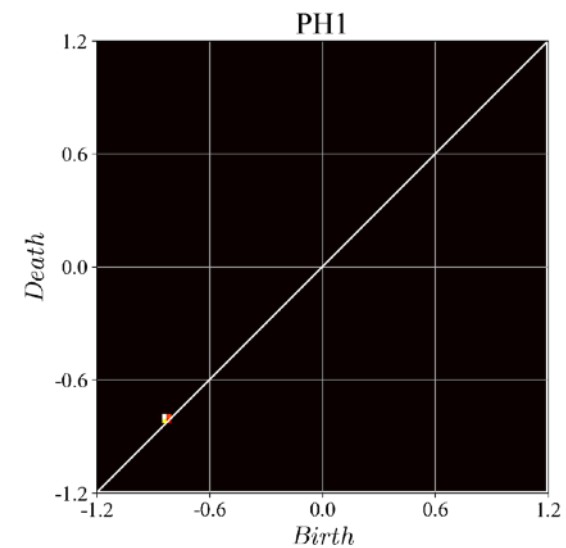
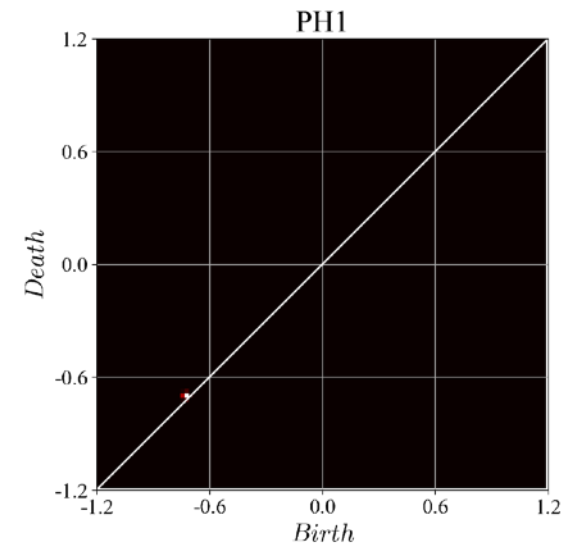
現象



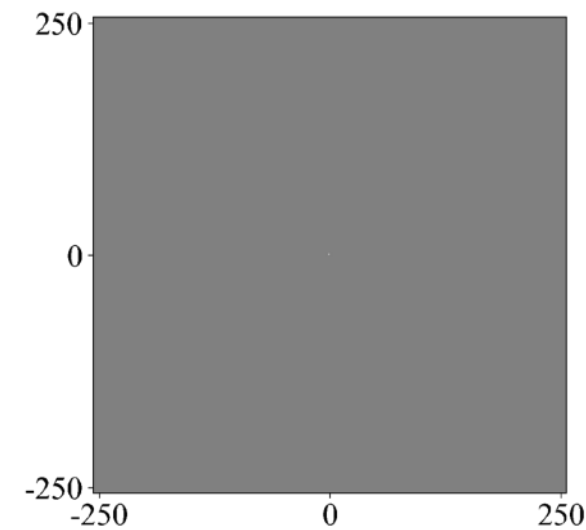
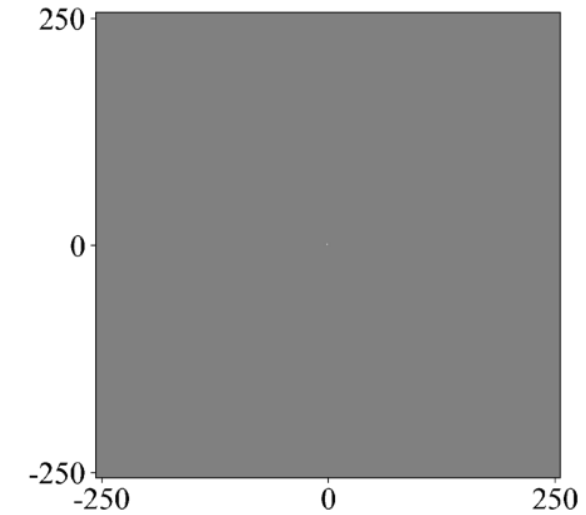
統計量



位相的特徴量



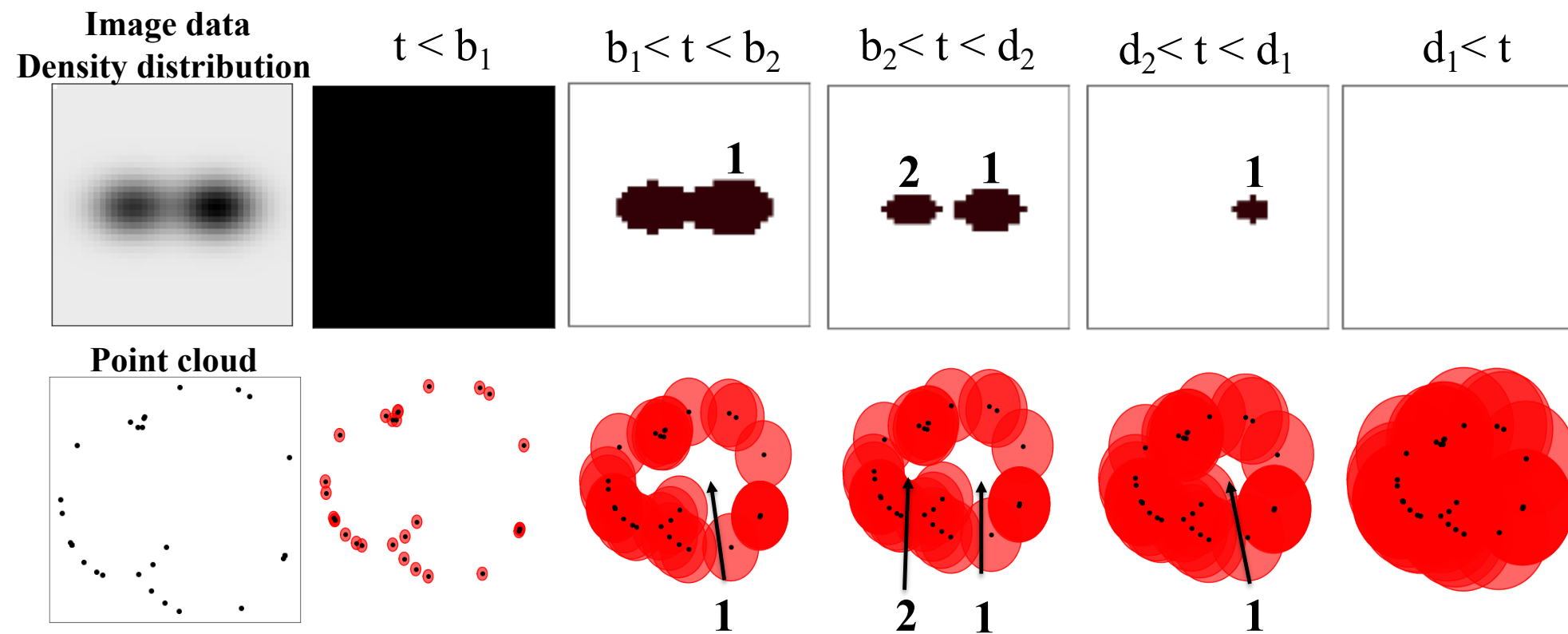
フーリエ基底



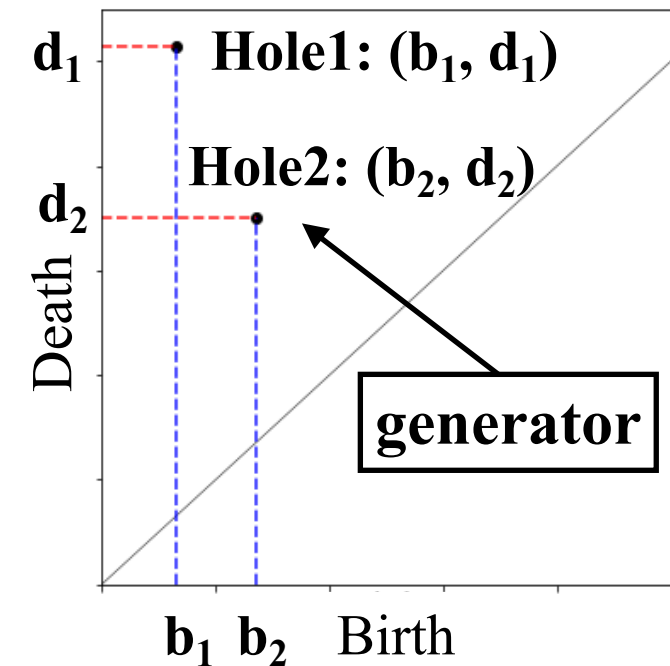
➡現象の違いに応じて見た目が大きく異なる特徴量となっている（[要件1]）



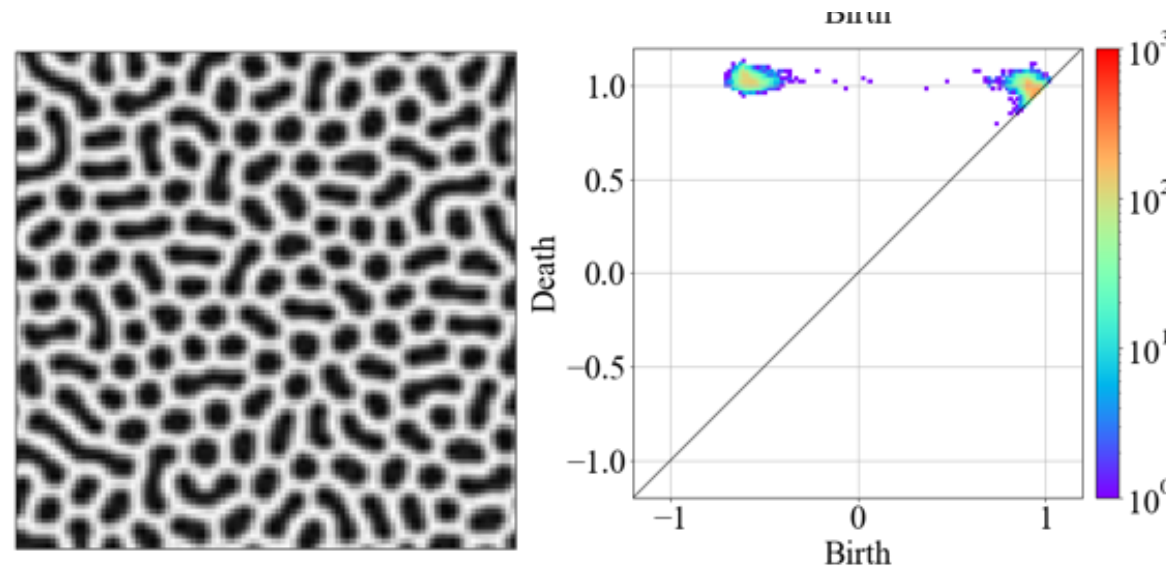
## パーシステントホモロジーの概説の概説

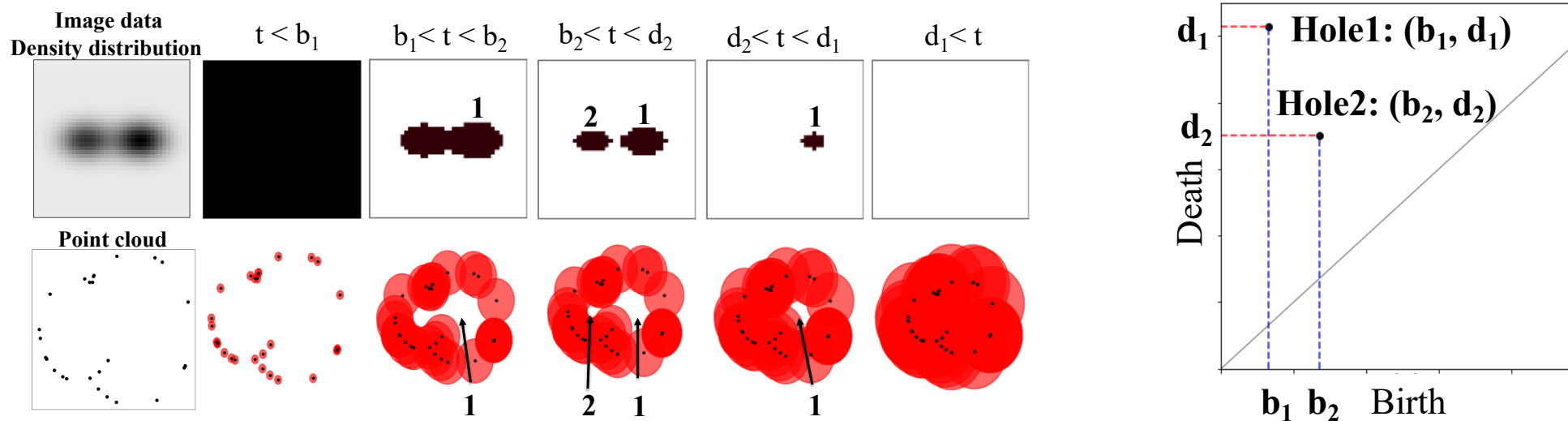


Persistence diagram (PD)



例：ある磁区構造におけるPD (のヒストグラム)





1. ひとつひとつの穴についての図形の増大列の情報を含む  
→ 閾値やスケールといったパラメータに応じて変化するトポロジカルな情報を取得できる
2. 逆解析が可能である (**[要件2]**)  
→ 特徴量空間と元の空間構造の対応関係がわかる
3. 安定性定理が成立する (**[要件3]**)  
→ 入力データの微小な変化に対して安定に振る舞う
4. 計算が容易

## 逆解析：消滅方体の中心位置

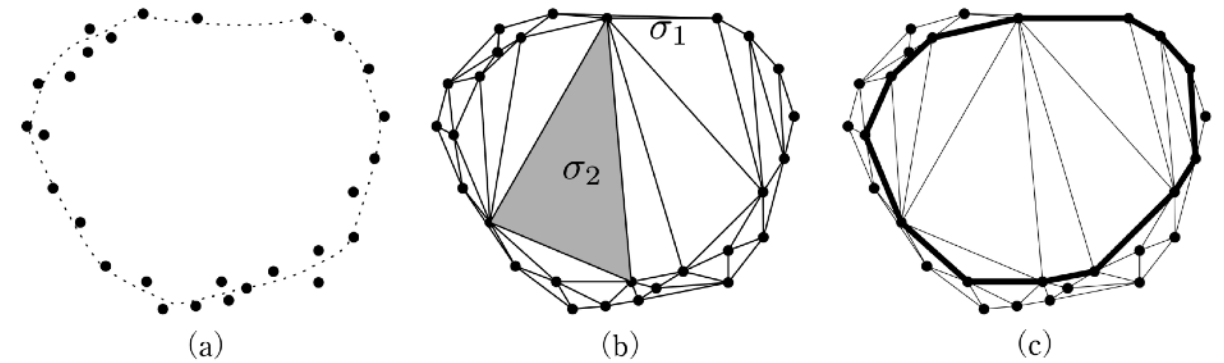
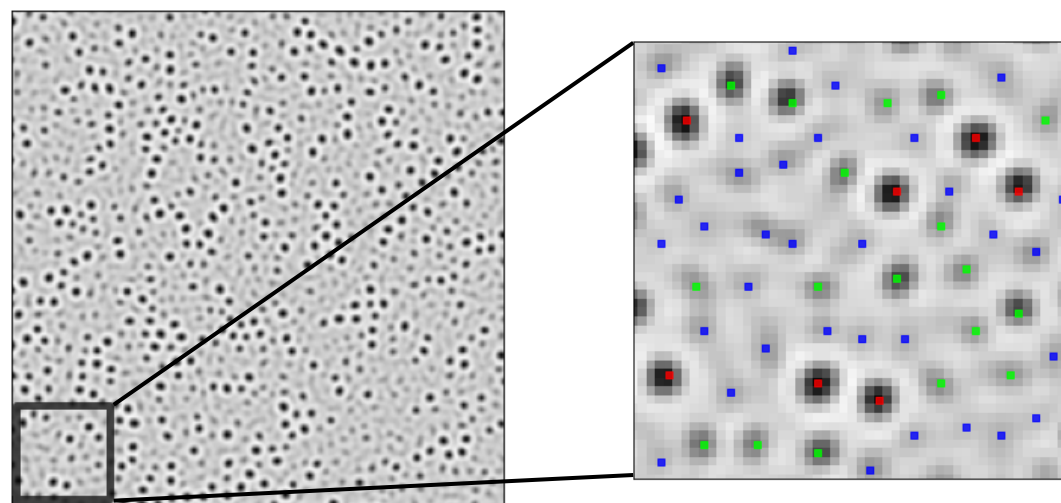
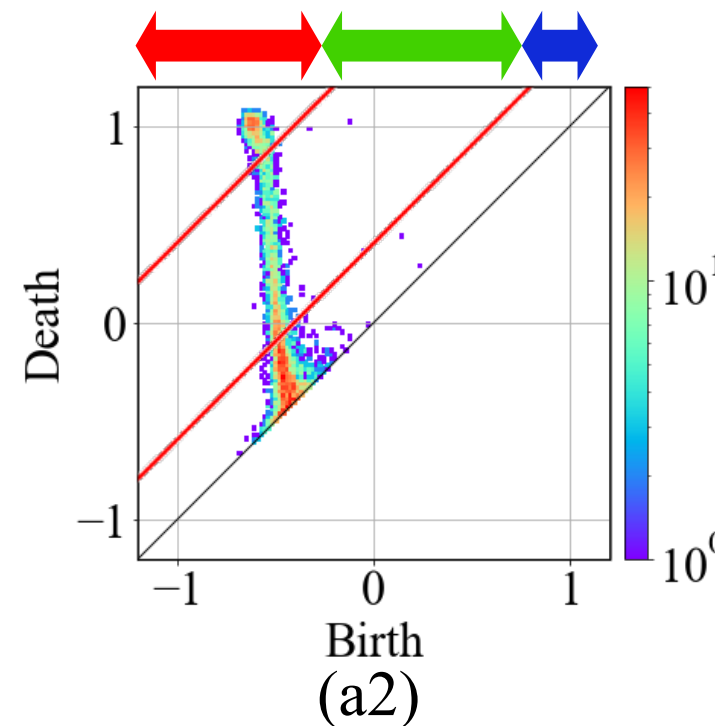


図8 発生/消滅単体と最小サイクル

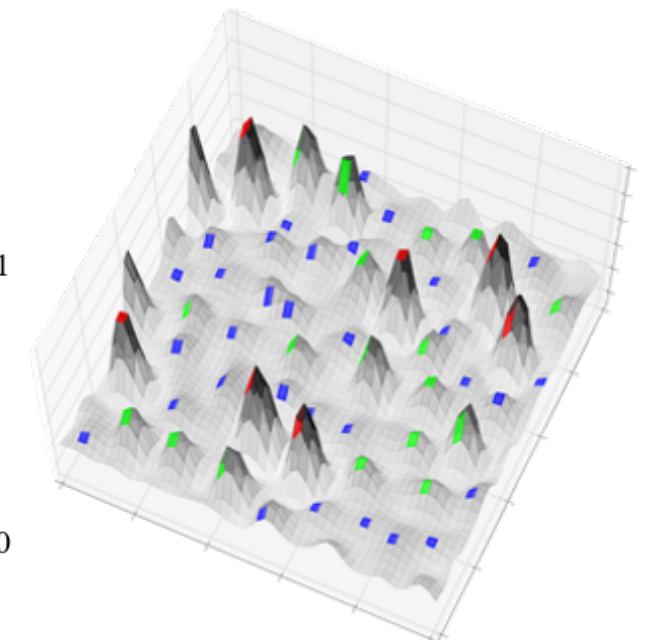
[大林一平, “パーシステンス図の逆問題,” 応用数理, **26-4** (2016), 151-158.]



(a1)

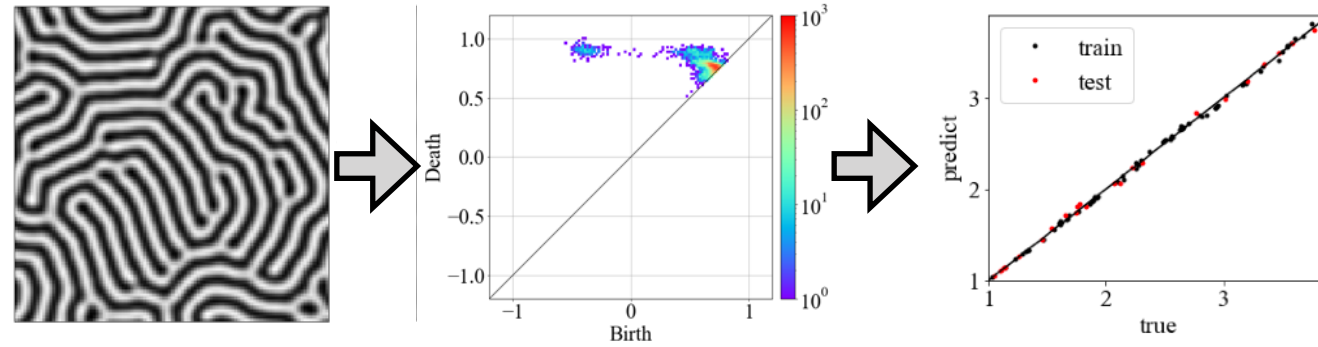


(a2)



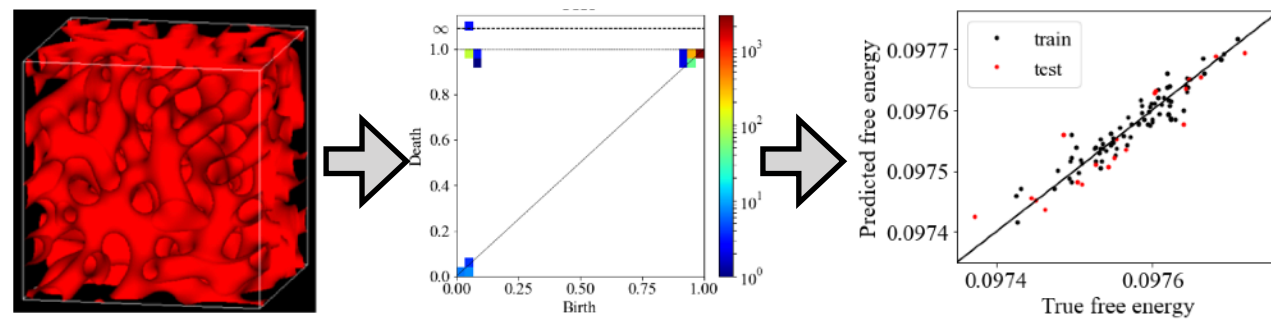
(a3)

## 強磁性体系のエネルギー支配因子の高精度推定

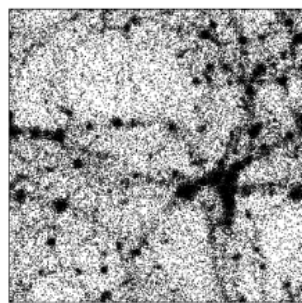


[Y. Mototake, M. Mizumaki, K. Kudo, K. Fukumizu, arxiv:2204.12194. (in submit).]  
[本武 陽一, 水牧 仁一朗, 工藤 和恵, 福水 健次, スマートプロセス学会誌, 2021]

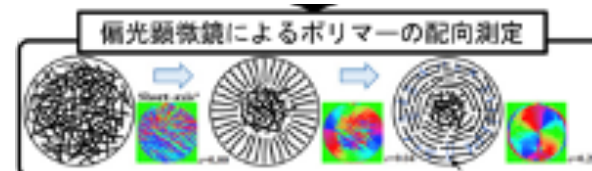
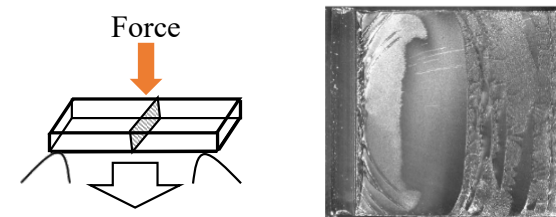
## 高分子ポリマー系の自由エネルギーの高精度推定



[Y. Mototake, S. Yamanaka, T. Aoyagi, T. Ohnishi, K. Fukumizu, NOLTA2020, 517 - 520, 2020]  
[Y. Mototake, S. Yamanaka, T. Aoyagi, T. Ohnishi, K. Fukumizu, J. Comp. Chem., Japan, 2021]



宇宙の  
大規模構造



破壊現象

磁性コロイド

磁性乱流・惑星物理

アクティブマター (群れ等)



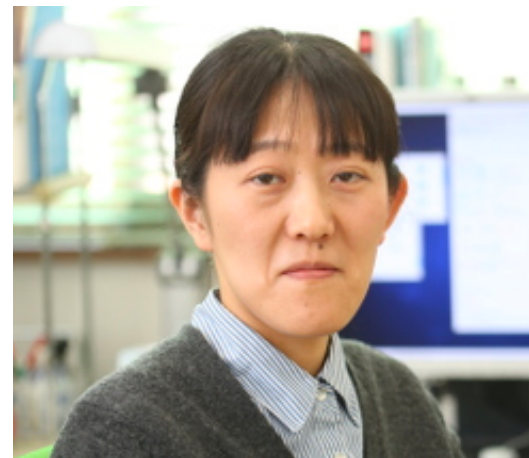
## Revealing the Mechanism of Magnetic Domain Formation by Topological Data Analysis

[(解説) 本武 陽一, 水牧 仁一朗, 工藤 和恵, 福水 健次, 「位相的データ分析法による材料構造形成過程の分析」, スマートプロセス学会誌, 2021]

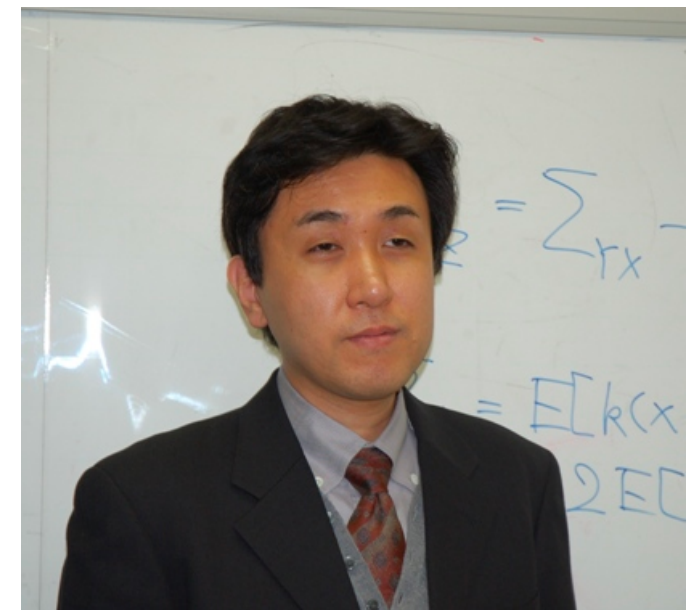
[Y. Mototake, M. Mizumaki, K. Kudo, K. Fukumizu, arXiv:2204.12194, (in submit)]



水牧 仁一朗 主幹研究員  
高輝度光科学研究センター  
(JASRI)



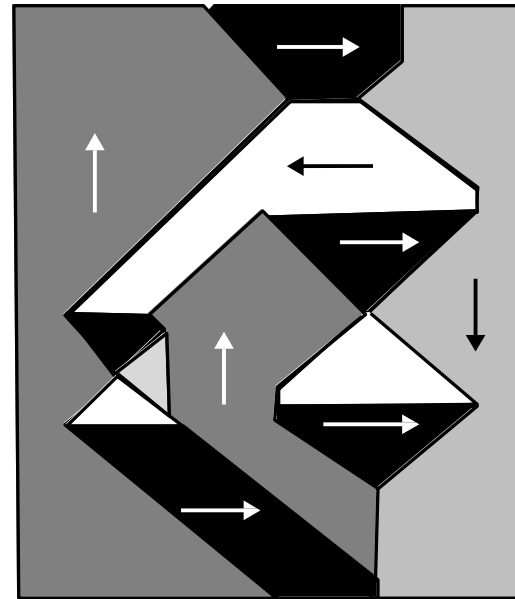
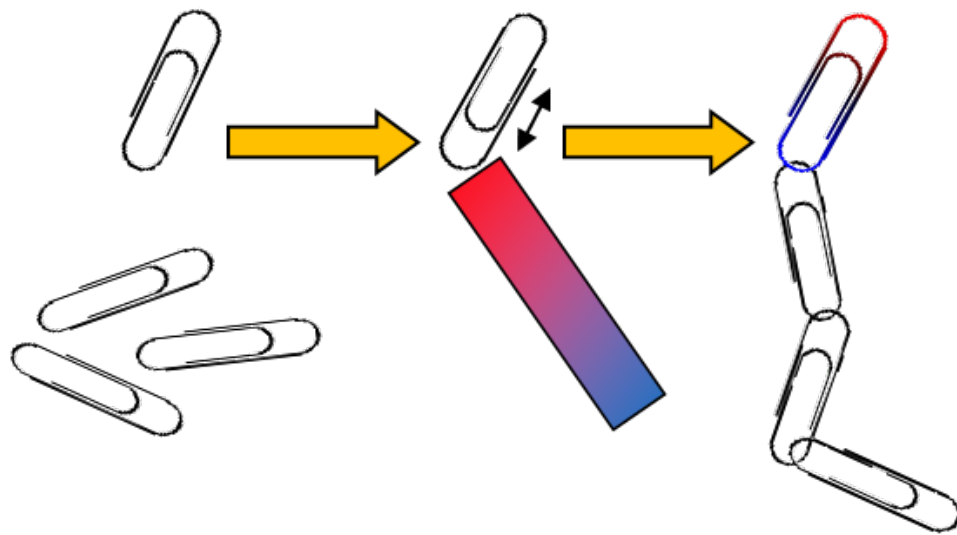
工藤和恵 准教授  
御茶ノ水女子大学  
東北大学



福水健次 教授  
統計数理研究所

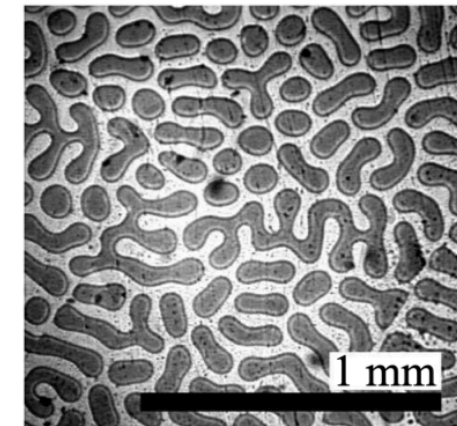


## 強磁性体の機序解明



Stripe

Labyrinth



Sea island structure Labyrinth structure

[Kazue Kudo, Michinobu Mino and Katsuhiro Nakamura 07]

[Katsuaki Sato, 「ようこそ、まぐねの国に」 より]

従来、永久磁石の磁区構造が磁気特性の重要な構成要素であることは理解されてきたが、磁区構造と、例えば「保磁力」が具体的にどのように相関するかの説明は「あいまい」である。

[小林 久理眞 教授「磁石の内部磁場分布とMFMおよびMOKEによる表面磁区構造の相関性」より引用]

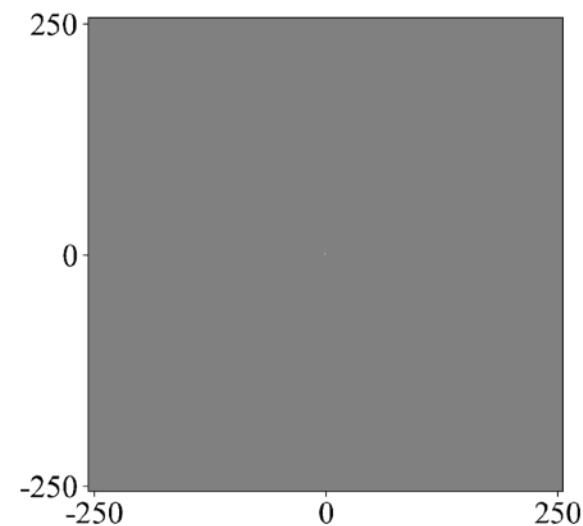
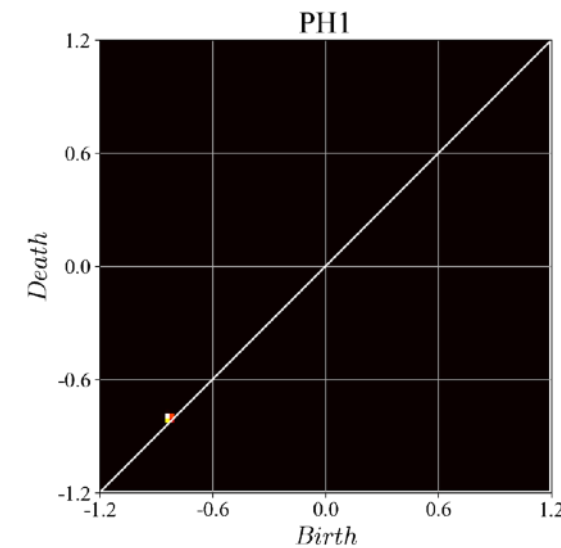
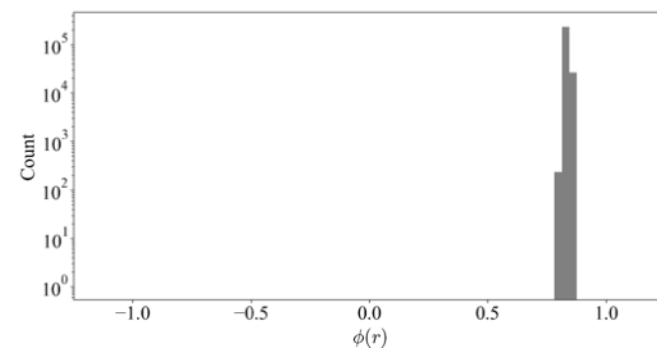
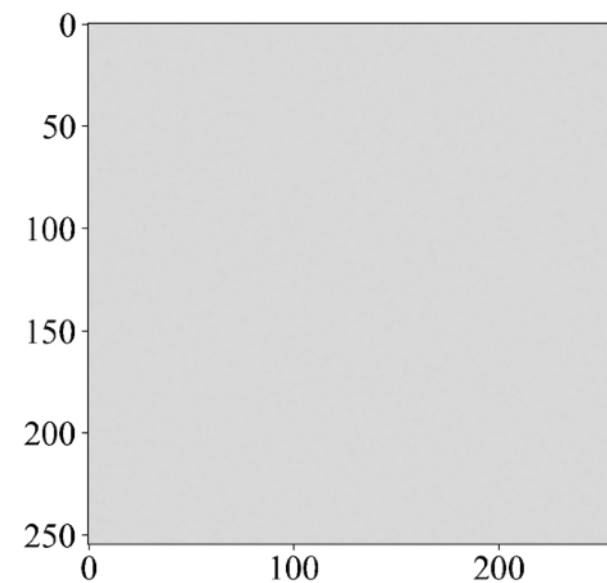
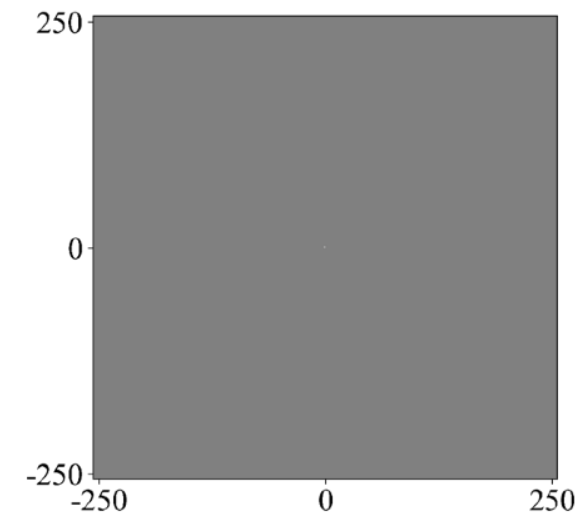
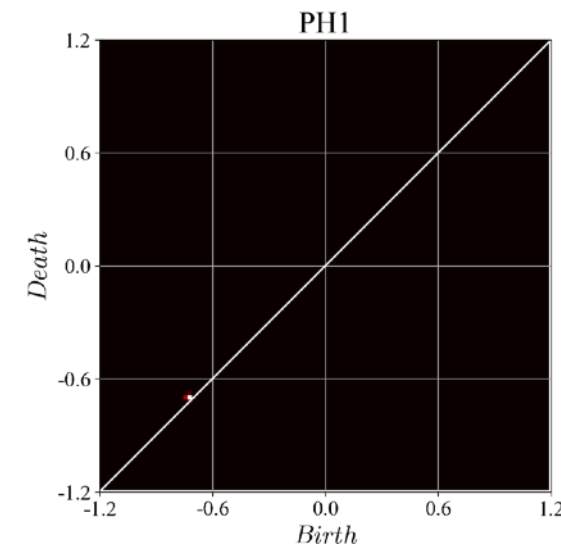
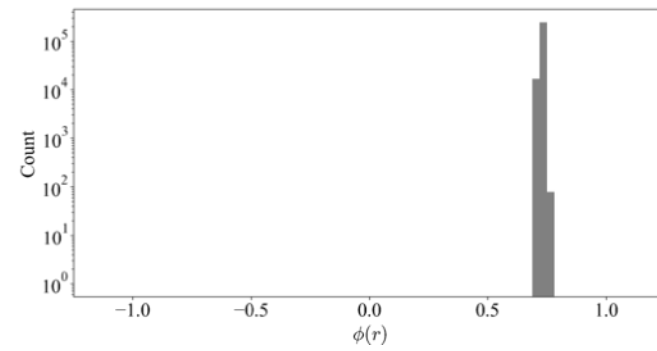
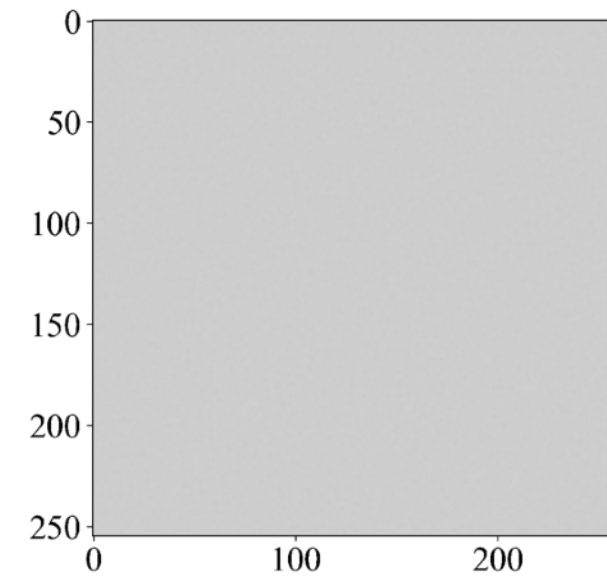
**▶磁区構造形成過程の位相的データ解析によって、磁区の幾何構造と物理機序との間の関係性の深い理解を促進する。**

現象

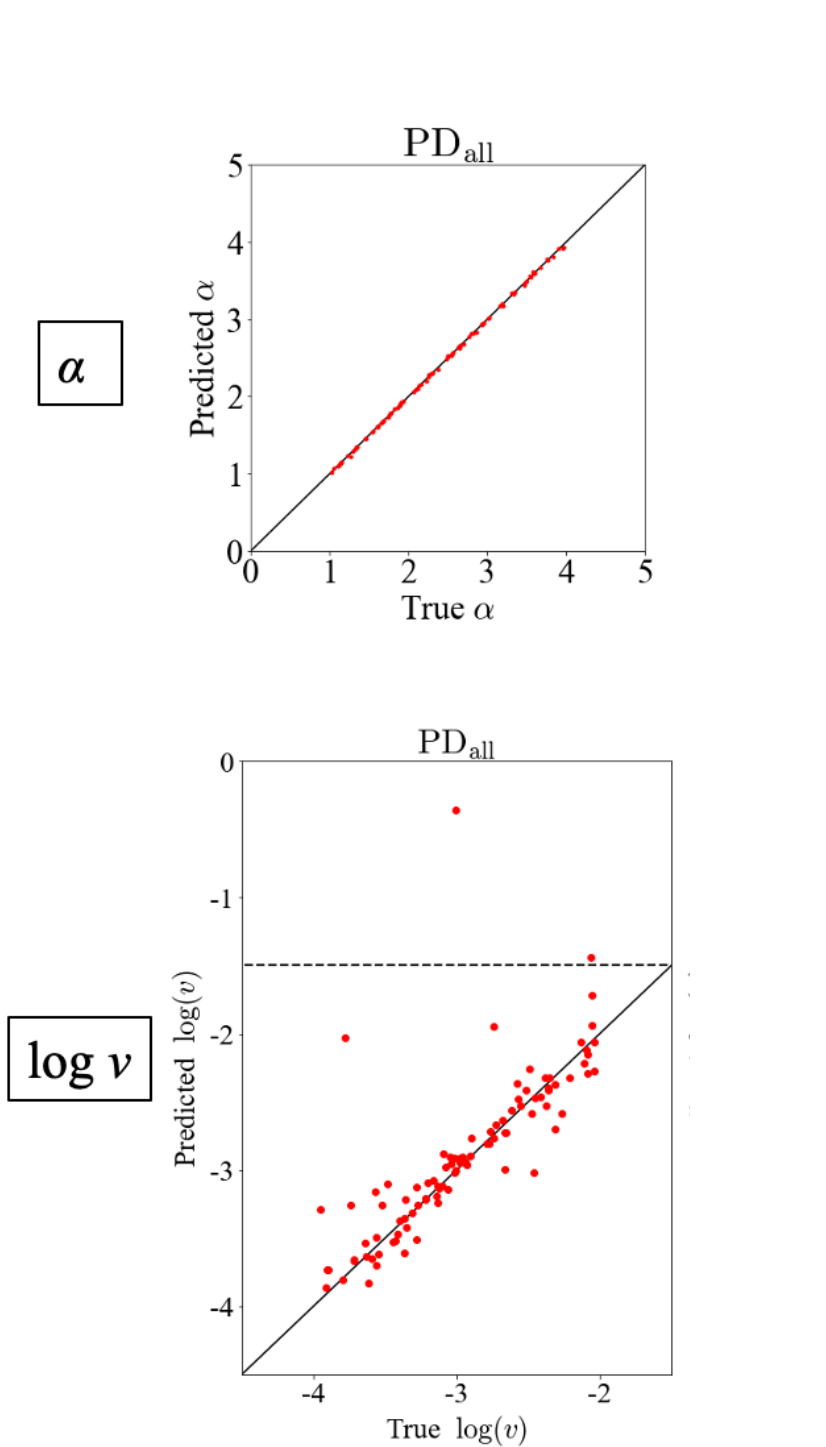
統計量

位相的特徴量

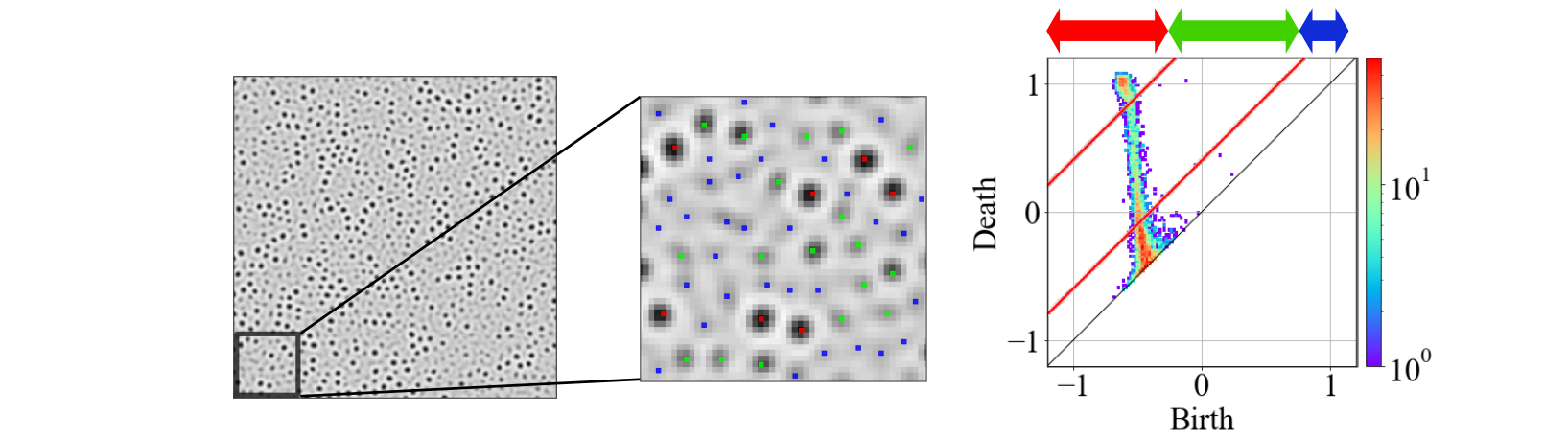
フーリエ基底



## 強磁性体系のエネルギー支配因子の推定

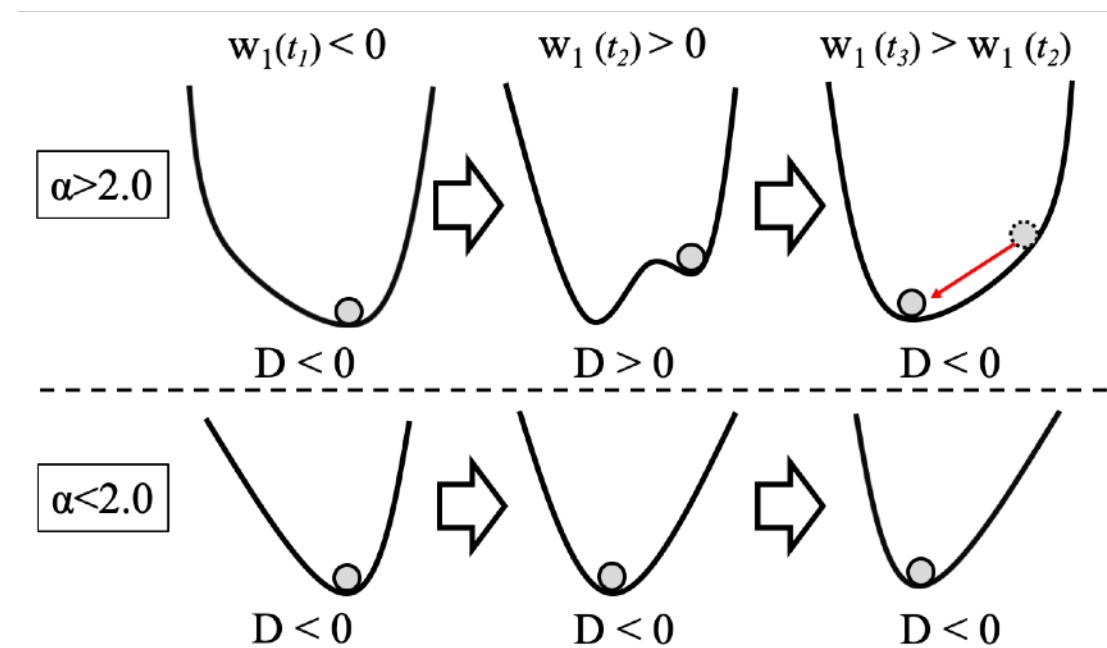


►系の支配因子を高精度に推定



$$E(\phi(\mathbf{r})) = \alpha \int d\mathbf{r} \lambda(\mathbf{r}) \left( -\frac{\phi(\mathbf{r})^2}{2} + \frac{\phi(\mathbf{r})^4}{4} \right) + \beta \int d\mathbf{r} \frac{|\nabla \phi(\mathbf{r})|}{2} + \gamma \int d\mathbf{r} d\mathbf{r}' \phi(\mathbf{r}) \phi(\mathbf{r}') G(\mathbf{r}, \mathbf{r}') + -h(t) \int d\mathbf{r} \phi(\mathbf{r})$$

$$E(\phi) = w_0(\beta) + w_1(\beta, \gamma, h(t))\phi + w_2(\alpha, \beta, \gamma)\phi^2 + w_4(\alpha)\phi^4$$

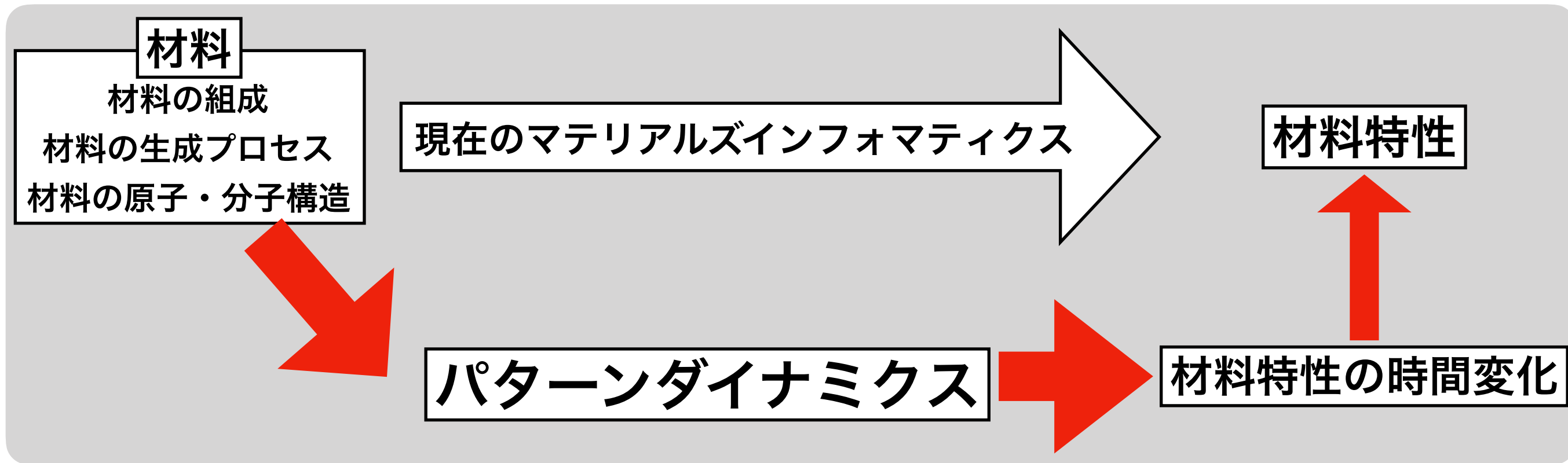


►逆解析を利用して系の物理的機序を表す縮約モデルを構築

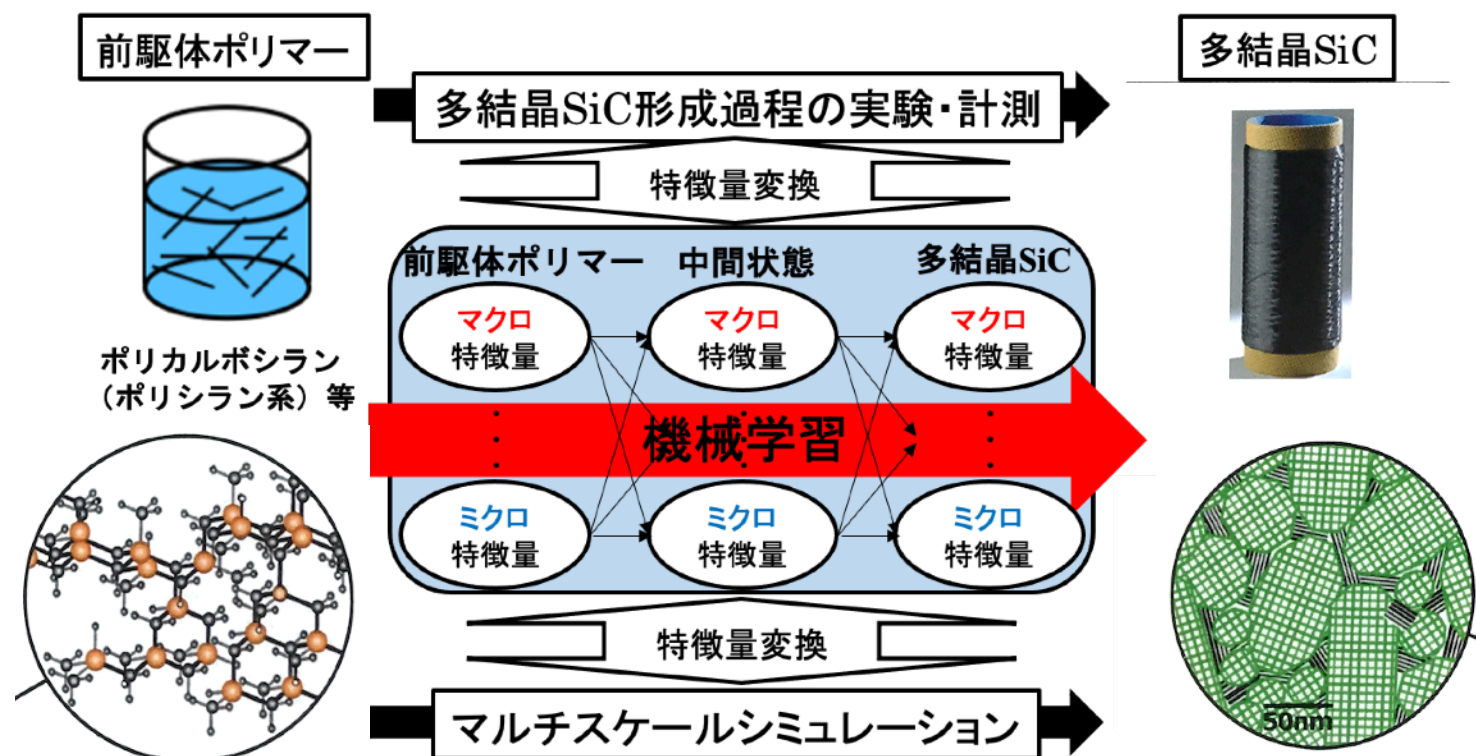
1. 自己紹介
2. 解釈可能AIとデータ駆動理学
3. DNN型アプローチによるデータ駆動理学の実現
4. 基底関数型アプローチによるデータ駆動理学の実現
5. **まとめと今後の展望**



「解釈可能AIによるパターンダイナミクスの数理構造抽出と材料情報学への応用」



「革新的セラミックス材料設計のための材料パターン情報学の創成」



[東京工科大学 香川研究室HPより]

[東京工科大学 香川研究室HPより]





## 機械学習による解釈可能AI構築アプローチ

