形状ベースクラスタリング手法を用いた 有限密度格子 Gross-Neveu 模型の相分類

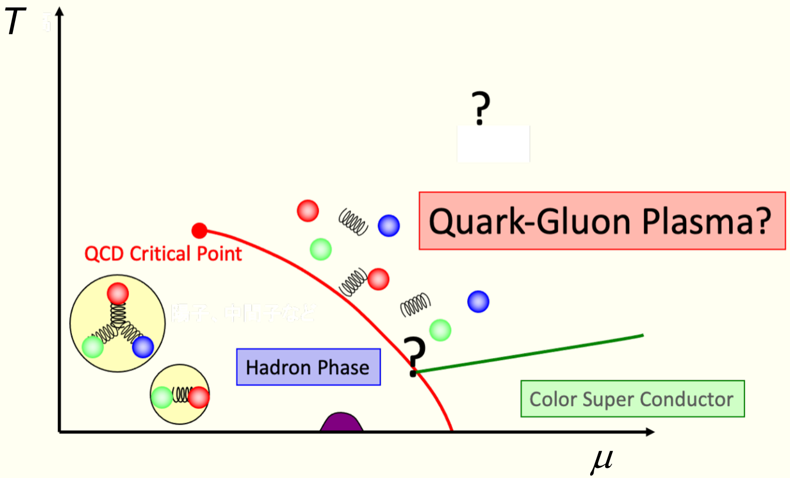
ChudenCTI Co.,Ltd.[A],
Graduate School of Advanced Science and Engineering, Hiroshima University[B],
International Institute for Sustainability with Knotted Chiral Metar Matter/SKCM[2],
Hiroshima University[C],
Kobayashi Maskawa Institute, Nagoya University[D]

Keita HORIE[A] and Chiho NONAKA[B,C,D]

CONTENTS

Motivation: Understanding of QCD Phase Diagram

QCD Phase Diagram



New method for classifying phases

Shape-based classifying method

Configurations of 1+1 d GN model on the lattice
Inhomogeneous phase,
homogeneously broken phase, chiral symmetric phase

Phases
Phase boundary

- At low density

  Lattice QCD
  High-energy heavy-ion collisions

- At high density

  Interesting phases:
  Inhomogeneous phase

  X Lattice QCD
  experiment ?

Motivation: (1+1)-dimensional Gross-Neveu Model

❖ Lagrangian density

$$\mathcal{L} = \bar{\psi} i \gamma^\nu \partial_\nu \psi + \frac{g^2}{2N} \left(\bar{\psi}\psi \right)^2 \qquad \sigma \sim \langle \bar{\psi}\psi \rangle$$

D. J. Gross and A. Neveu, Phys. Rev. D 10, 3235 (1974)

Important features from comparison with QCD

- ❑ Asymptotic freedom
- ❑ Spontaneous symmetry breaking of discrete chiral symmetry

$$\psi \to \gamma_5 \psi, \quad \bar{\psi} \to -\bar{\psi}\gamma_5$$

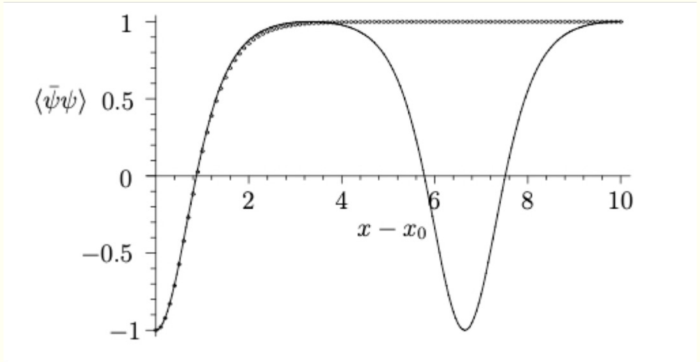
- ❑ No sign problem : Monte Carlo simulation
- ❑ Inhomogeneous chiral condensate in large N_f limit

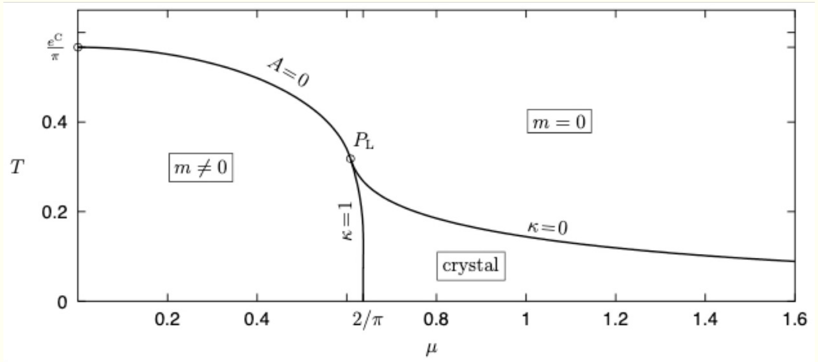
Motivation: (1+1)-dimensional GN Model @ Continuous Theory

❖ Lagrangian density in the continuous theory

$$\mathcal{L} = \bar{\psi} i \gamma^{\nu} \partial_{\nu} \psi + \frac{g^2}{2N} \left(\bar{\psi} \psi \right)^2 \qquad \sigma \sim \langle \bar{\psi} \psi \rangle$$ in the large N_f limit

❖ Specific ansatz & phase diagram





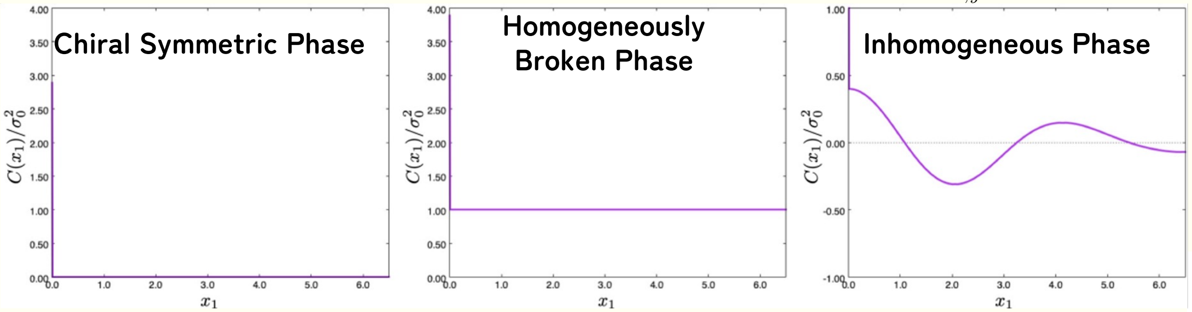
O. Schnetz, M. Thies and K. Urlichs, Annals Phys. 314, 425-447 (2004)

To calculation without ansatz, we use lattice field theory

J. Lenz et. al, Phys. Rev. D 101, no.9, 094512 (2020)

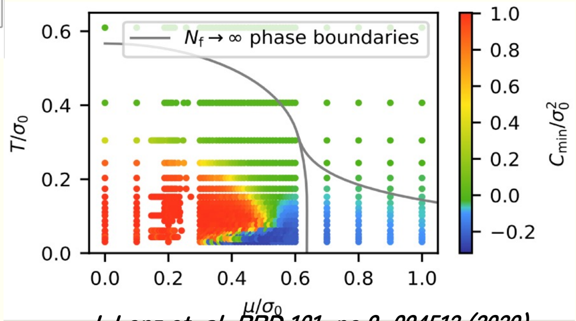
Motivation: (1+1)-dimensional GN Model @ Lattice Theory

- **Spatial correlators in three phases** $C(x) = \frac{1}{N_t N_x} \sum_{t,y} \langle \sigma(t, y+x) \sigma(t,y) \rangle$



- **Classification by minimum of the spatial correlator**

$$C_{\min} := \min_x C(x) \begin{cases} \gg 0, & \text{the Homogeneously Broken Phase} \\ \approx 0, & \text{the Chiral Symmetric Phase} \\ < 0, & \text{the Inhomogeneous Phase} \end{cases}$$



J. Lenz et. al, PRD 101, no.9, 094512 (2020)

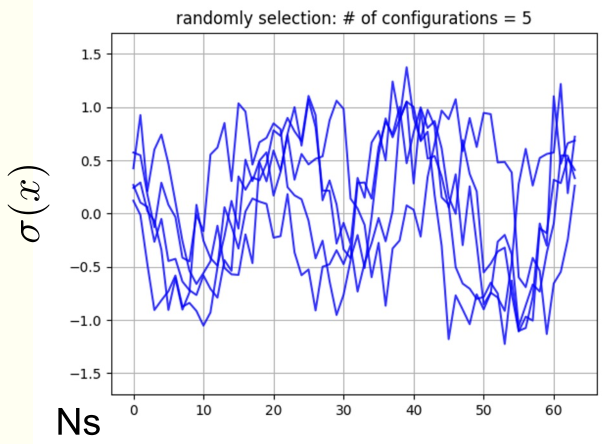
Knowledge of Ansatz is used in the interpretation of the results.

Can we directly extract the spatial dependence of the configurations $\sigma(x)$?

Method: Shape-based Clustering Method

❖ Difficulty in direct extraction of the spatial dependence

The configurations shifted at each Monte Carlo step



When we generate enough number of configurations, <u>the expectation value at each point becomes zero</u>.

⇨ spatial correlators

Is there a more general and direct method?

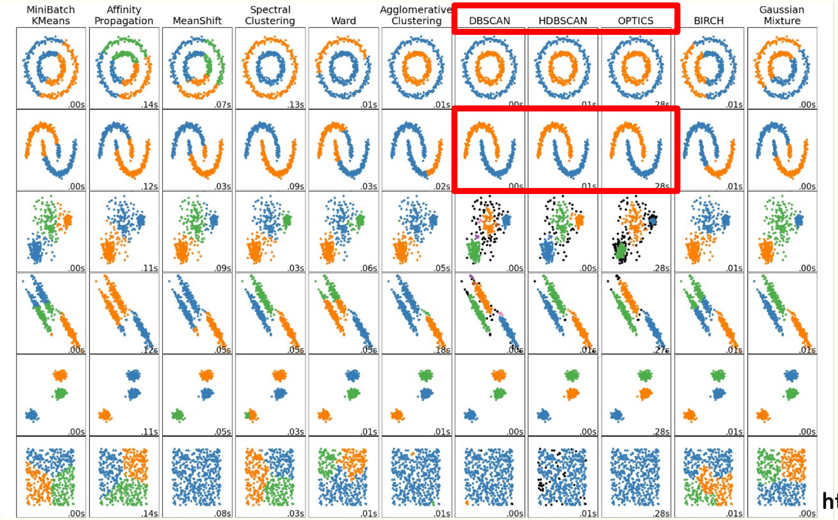
We need **shift-invariant** clustering method that focuses on the shape of configurations.

Time Series Clustering Method

Method: Shape-based Clustering Method

❖ Clustering Method

- ❑ **Unsupervised Learning**
  A method for grouping data without labeled training data
- ❑ It is important to choose a "similarity" that represents how similar two data points are.



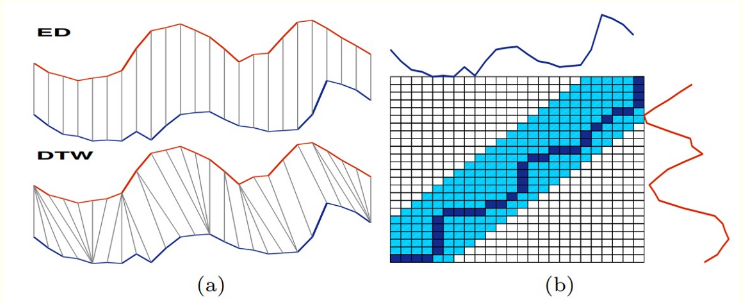
The choice of similarity can lead to differences in the data that can be grouped.

Example:
Moon-shaped data are well-suited for density-based similarity calculations such as DBSCAN.

https://scikit-learn.org/stable/modules/clustering.html

解析手法：形状ベースクラスタリング手法の応用

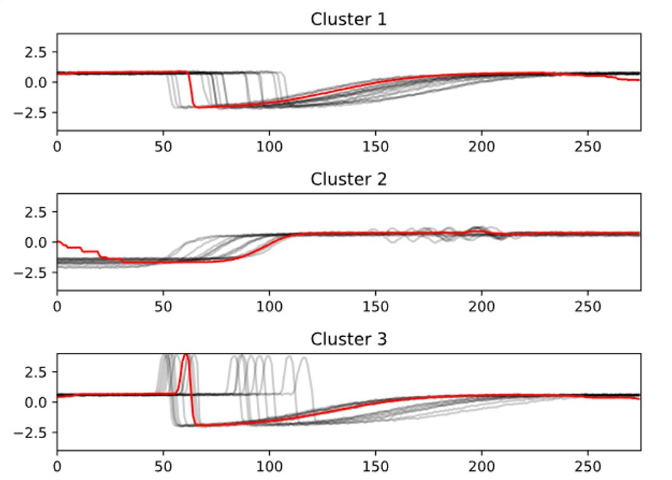
❖ K-Shape法
- ➢ 教師なし学習手法の一種
- ➢ 時系列データのクラスタリング手法
- ➢ シフト不変性とスケール不変性の特徴を持ったアルゴリズムを使用



J. Paparrizos, L. Gravano, PROC. ACM SIGMOD Int. Conf. Manage. Data, pp. 1855-1870, 2015

❖ K-Shape法の特徴
- ➢ DTW法よりも高速
  ※DTW法もシフト不変性とスケール不変性を持つ
- ➢ インプット　：時系列データ, クラスタ分割数, etc…
  アウトプット：クラスタリング結果, クラスタ重心

右図の赤い実線が
クラスタ重心（クラスタ代表値）



R. Tavenard et. al., JMLR, 21, no.118 (2020)
http://jmlr.org/papers/v21/20-091.html

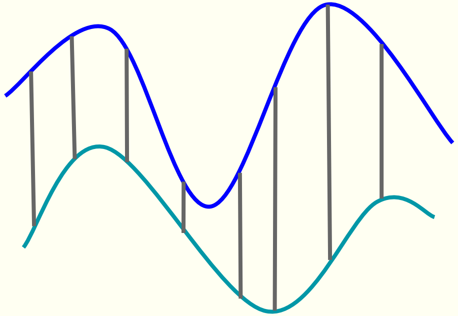
Method: Shape-based Clustering Method

❖ Time Series Clustering Method

- ❑ One of the Clustering methods
- ❑ Clustering methods for time series (1-dimensional data):
- ❑ Major methods:

  Dynamic Time Wraping, K-Shape, k-means, etc...

❖ Simple Example: k-means method with Euclidean distance



“Similarity”:

$$d_E(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{(\boldsymbol{x} - \boldsymbol{y})^2}$$

Note：
k-means method is not well suited for time series data because the similarity decreases when the phase is shifted.

Method: Shape-based Clustering Method

- K-Shape Method

This method is characterized by shift-invariance and scale-invariance

- Shape-based Distance (SBD)：the similarity with shift-invariance

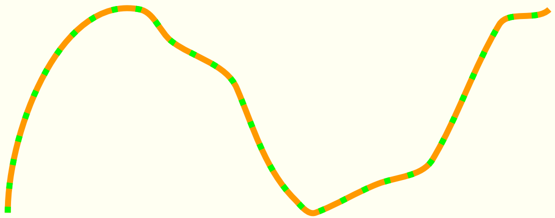
$$SBD(\sigma_1(\boldsymbol{x}), \sigma_2(\boldsymbol{x})) = 1 - \max_{w}\left(\frac{CCF_w(\sigma_1(\boldsymbol{x}), \sigma_2(\boldsymbol{x}))}{\sqrt{ACF(\sigma_1(\boldsymbol{x}))ACF(\sigma_2(\boldsymbol{x}))}}\right)$$

CCF：cross-correlation function

ACF：auto-correlation function

J. Paparrizos, L. Gravano, PROC. ACM SIGMOD Int. Conf. Manage. Data, pp. 1855-1870, 2015

SBD evaluates similarity ignoring the phase shift.



Method: Shape-based Clustering Method

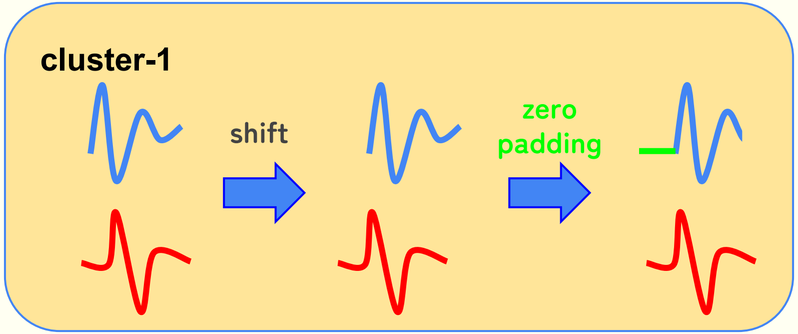
❖ Algorithm - Refinement step

INPUT
X is an *n*-by-*m* matrix containing *n* time series of length *m* that are initially z-normalized.
k is the number of clusters to produce.

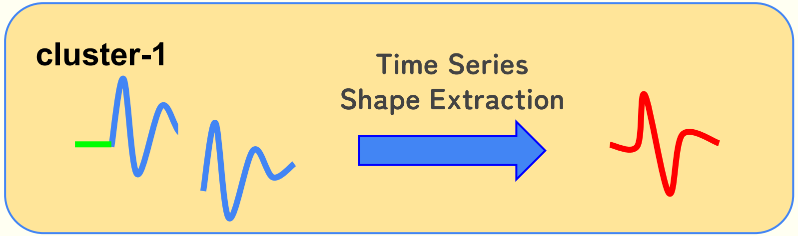
while cluster labels don't change or *iter* < max
|
| Refinement step
|
| Assignment step
L

OUTPUT
IDX is as *n*-by-1 vector containing the assignment of *n* time series to *k* clusters (initialized randomly).
C is a *k*-by-*m* matrix containing *k* centroids of length *m* (initialized as vectors with all zeros).

1. Shift each data to overlap with the centroid with minimized SBD



2. Calculate the optimal centroid for each data in the cluster (Time Series Shape Extraction)



Method: Shape-based Clustering Method

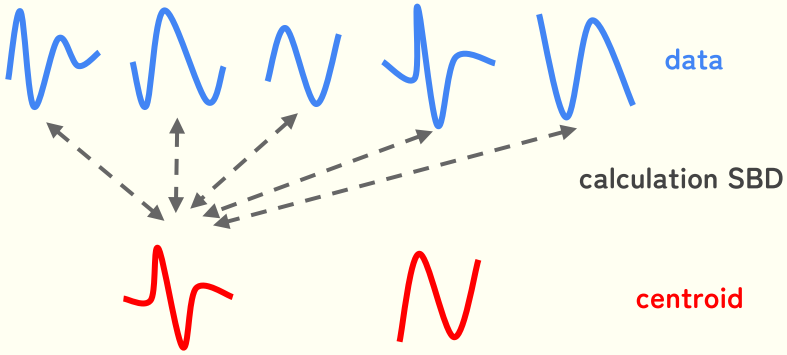
❖ Algorithm - Assignment step

INPUT
X is an *n*-by-*m* matrix containing *n* time series of length *m* that are initially z-normalized.
k is the number of clusters to produce.

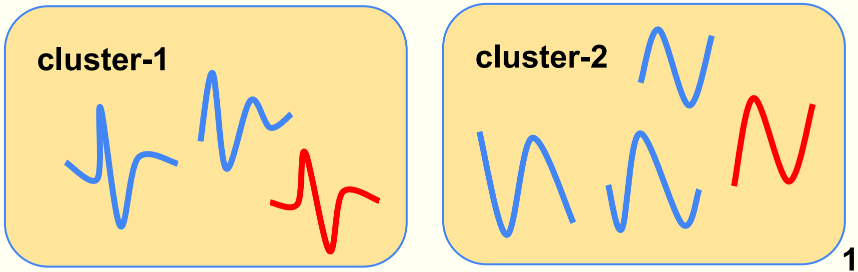
while cluster labels don't change or *iter* < max
|
| Refinement step
|
| Assignment step
L

OUTPUT
IDX is as *n*-by-1 vector containing the assignment of *n* time series to *k* clusters (initialized randomly).
C is a *k*-by-*m* matrix containing *k* centroids of length *m* (initialized as vectors with all zeros).

1. Calculate the similarity (SBD) between each data and the centroid



2. Assign the centroid with the maximum similarity to each data

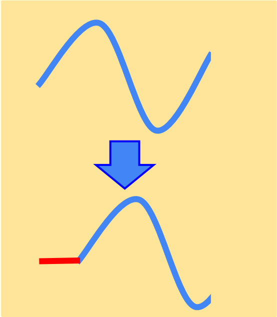


Method: Shape-based Clustering Method

- Modify K-Shape for lattice simulation

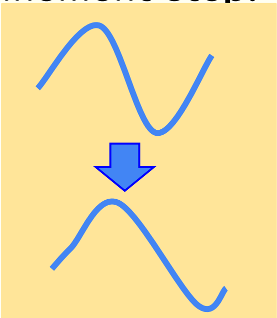
The Original K-Shape：we peform zero padding in the refinement step.

$$\boldsymbol{x} = \begin{cases} \overbrace{0,\ldots,0}^{s}, x_1, x_2, \ldots, x_{m-s}, & s \geq 0 \\ x_{1-s}, \ldots, x_{m-1}, x_m, \underbrace{0,\ldots,0}_{|s|} & s < 0 \end{cases}$$



The Modified K-Shape：we impose periodic boundary condition in the refinement step.

$$\boldsymbol{x} = \begin{cases} x_{m-s+1}, \ldots, x_m, x_1, x_2, \ldots, x_{m-s}, & s \geq 0 \\ x_{1-s}, \ldots, x_{m-1}, x_m, x_{m+1}, \ldots, x_{-s} & s < 0 \end{cases}$$



Results

❖ Lattice Simulation Setup

- ➢ We use a standard hybrid Monte Carlo algorithm.
- ➢ Lattice discretization of fermions is a naive fermion.
- ➢ To set the scale, we use the expectation value at zero temperature and zero chemical potential.
- ➢ We use the same coupling constant as the previous study [J. Lenz et. al, PRD 101, no.9, 094512 (2020)].
- ➢ The other simulation parameters are described in the table below:

fermion	N_f	$N_s = L/a$	$N_t = 1/Ta$	g^2	$a\sigma_0$	μ/σ_0
naive	8	64	14, 24, 64	1.8132	0.42 ± 0.01	0.0, 0.5, 0.6

❖ Preprocessing

In the equilibrium, configurations do not have time dependence.

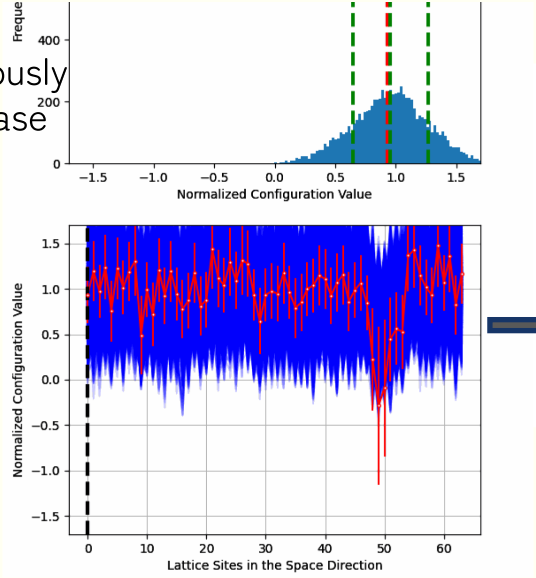
Therefore, we calculate the average of configurations along the time axis.

Results: Shape-based Clustering Method

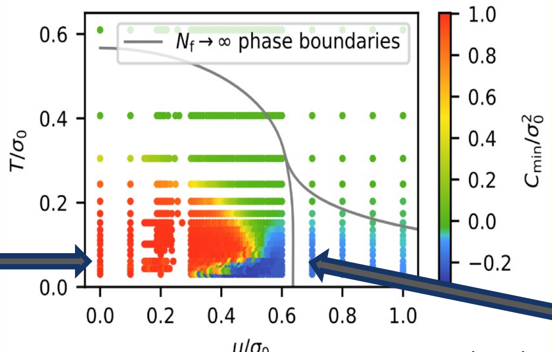
❖ **Extraction the spatial dependence of the configurations**

➢ **Set the cluster number k = 1**

homogeneously broken phase

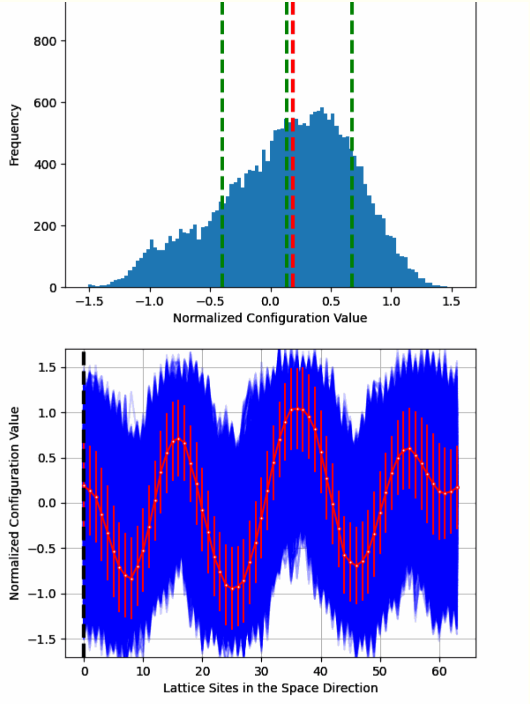


$$(T/\sigma_0, \mu/\sigma_0) = (0.037, 0.000)$$



J. Lenz et. al, PRD 101, no.9, 094512 (2020)

inhomogeneous phase



$$(T/\sigma_0, \mu/\sigma_0) = (0.098, 0.630)$$

We can extract the spatial dependence without the ansatz!

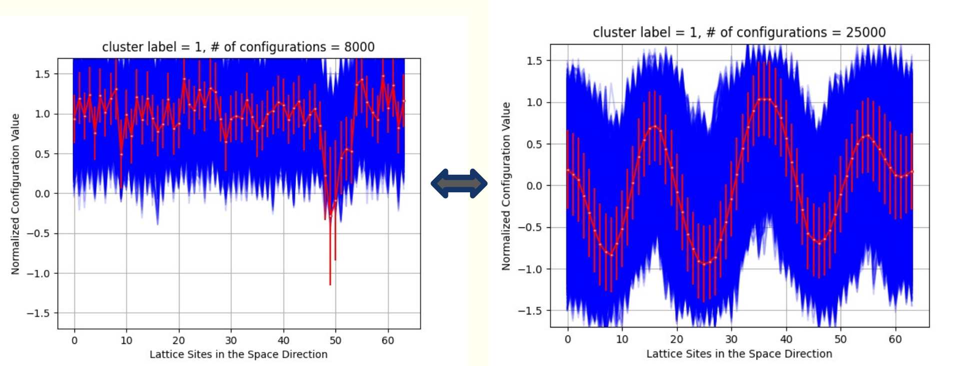
Summary:

❖ Summary

- ➢ We applied a shape-based clustering method, a type of unsupervised learning, to the analysis of lattice configurations.
- ➢ We modified the method to make it suitable for lattice calculations, including periodic boundary condition.

$$\boldsymbol{x} = \begin{cases} \overbrace{0,\ldots,0}^{s}, x_1, x_2, \ldots, x_{m-s}, & s \geq 0 \\ x_{1-s}, \ldots, x_{m-1}, x_m, \underbrace{0,\ldots,0}_{|s|} & s < 0 \end{cases} \Rightarrow \boldsymbol{x} = \begin{cases} x_{m-s+1}, \ldots, x_m, x_1, x_2, \ldots, x_{m-s}, & s \geq 0 \\ x_{1-s}, \ldots, x_{m-1}, x_m, x_{m+1}, \ldots, x_{-s} & s < 0 \end{cases}$$

- ➢ We succeeded in extracting the spatial dependence without the knowledge of ansatz.



❖ Future Work

- ➢ We will apply this method to the configurations to classify the phases of GN model.
- ➢ We will use this method to other models with spatial dependent phases.