Interpretable and unsupervised phase classification

Julian Arnold, Frank Schäfer, Martin Žonda, Axel U. J. Lode

University of Basel

Bruder group

General framework

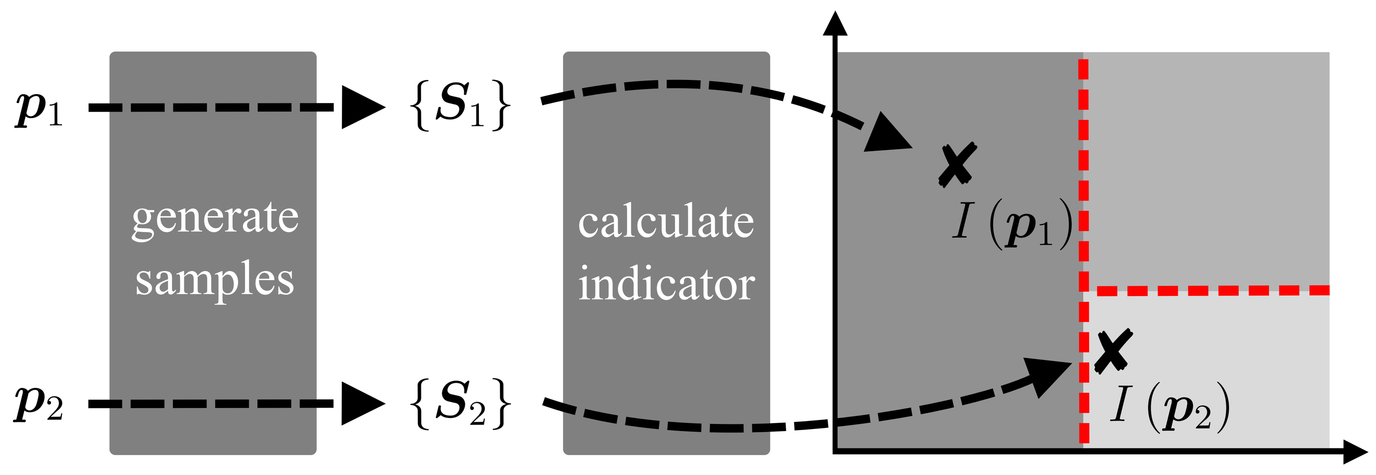
SETUP: The parameter space of a physical system is sampled at a set of points $\{\boldsymbol{p}_i\}$

- at each point $\boldsymbol{p}_i$, samples $\{\boldsymbol{S}_i\}$ that represent the system's state are generated

TASK: Automated construction of the phase diagram in an

- unsupervised way ⇔ without prior knowledge of phases
- interpretable way ⇔ without black-box models

IDEA: **CALCULATE INDICATOR $I(\boldsymbol{p})$ THAT HIGHLIGHTS PHASE BOUNDARIES**



Mean-based method

$$I(\boldsymbol{p}) = \sum_i \left\| \frac{\partial \bar{\boldsymbol{x}}}{\partial p_i} \right\| \approx \sum_i \frac{\| \bar{\boldsymbol{x}}(\boldsymbol{p}+\boldsymbol{e}_i \Delta p_i) - \bar{\boldsymbol{x}}(\boldsymbol{p}-\boldsymbol{e}_i \Delta p_i)\|}{2\Delta p_i},$$

where $\bar{\boldsymbol{x}}(\boldsymbol{p})$ is the mean input at $\boldsymbol{p}$

- representation $\mathcal{R}: \boldsymbol{S} \to \boldsymbol{x}$, HERE: correlation functions measuring square, axial, and diagonal correlations

Prediction-based method

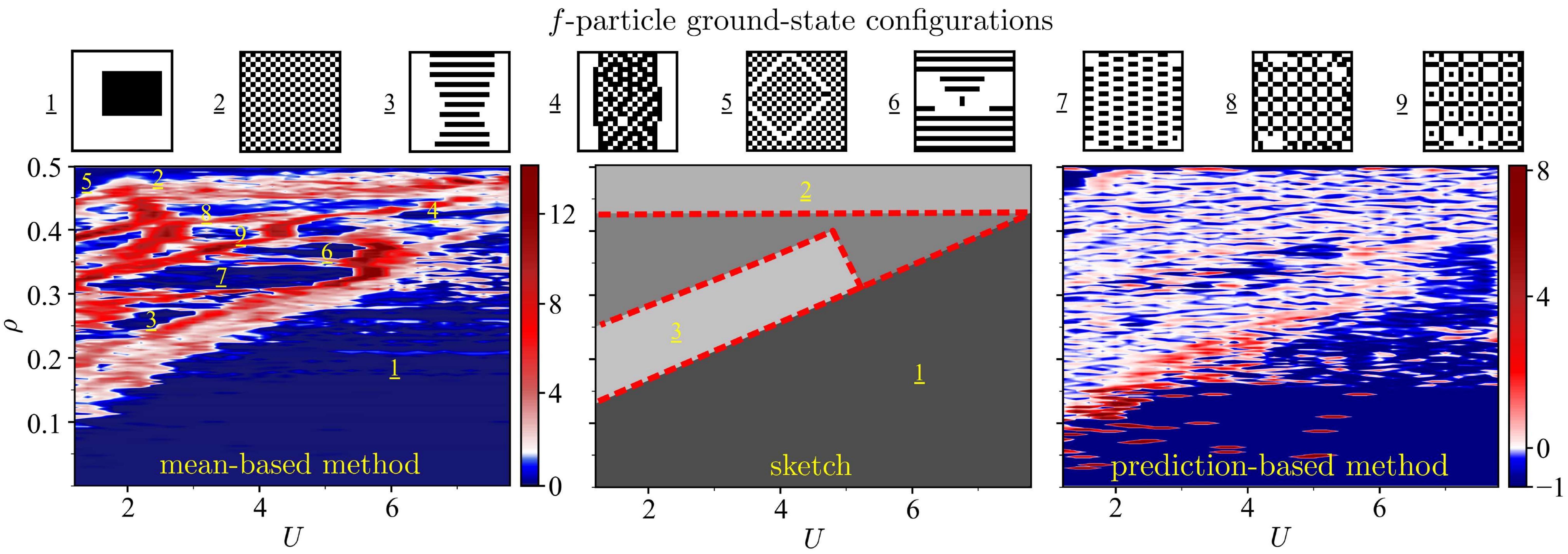
$$I(\boldsymbol{p}) = \nabla_{\boldsymbol{p}} \cdot \boldsymbol{\delta p} \approx \sum_i \frac{\delta p_i(\boldsymbol{p}+\boldsymbol{e}_i \Delta p_i) - \delta p_i(\boldsymbol{p}-\boldsymbol{e}_i \Delta p_i)}{2\Delta p_i},$$

where $\boldsymbol{\delta p} = \hat{\boldsymbol{p}} - \boldsymbol{p}$ is the deviation of prediction $\hat{\boldsymbol{p}}$ from underlying system parameter $\boldsymbol{p}$

- predictive model $m: \boldsymbol{x} \to \hat{\boldsymbol{p}}$ trained on a mean-square-error loss function, HERE: deep neural network

Ground-state phase diagram of the Falicov-Kimball model

The Falicov-Kimball model describes mobile d- and localized f-particles on a lattice: $\hat{H} = -t\sum_{\langle ij \rangle}(\hat{d}_i^\dagger \hat{d}_j + \hat{d}_j^\dagger \hat{d}_i) + U\sum_i \hat{d}_i^\dagger \hat{d}_i \hat{f}_i^\dagger \hat{f}_i$



- simple analytical expression for indicator ⇒ directly interpretable
  - allows for direct physical insight into the revealed phases and phase transitions
- indicator measures magnitude of change in mean ordering across phase boundary

- relies on a highly expressive predictive model ⇒ *not* directly interpretable
  - rendered interpretable via derivation of analytical expression for optimal indicator $I_{\text{opt}}(\boldsymbol{p})$ based on predictive model m_{opt} that minimizes loss function
- indicator measures mean extent of neighbouring phases in parameter space

⇒ TWO DISTINCT APPROACHES THAT PROVIDE COMPLEMENTARY INSIGHTS INTO THE PHASE DIAGRAM

References

[F. Schäfer and N. Lörch, Phys. Rev. E **99**, 062107 (2019)]
[E. Greplova *et al.*, New J. Phys. **22**, 045003 (2020)]
[J. Arnold *et al.*, Phys. Rev. Res. **3**, 033052 (2021)]

Contacts

julian.arnold@unibas.ch
frank.schaefer@unibas.ch

Acknowledgments

FNSNF Swiss National Science Foundation
QSIT Quantum Science and Technology — National Centre of Competence in Research