Data-driven discovery of relevant information in many-body problems

Roberto Verdel Aranda
ICTP, Trieste

“Machine learning for lattice field theory and beyond”
ECT*, Trento
29/06/2023

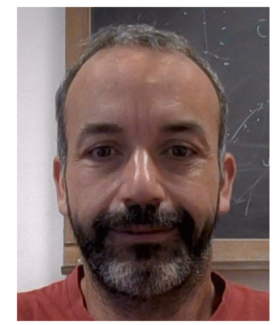
The Abdus Salam International Centre for Theoretical Physics

ECT* EUROPEAN CENTRE FOR THEORETICAL STUDIES IN NUCLEAR PHYSICS AND RELATED AREAS

Collaborators

R. K. Panda
(ICTP/SISSA)

V. Vitale
(U. Grenoble Alps)



A. Rodriguez
(UniTS)

M. Dalmonte
(ICTP/SISSA)

S. Pedrielli
(UniTS -> TU Berlin)

E. Donkor
(ICTP/SISSA)

H. Sun
(QMU London)

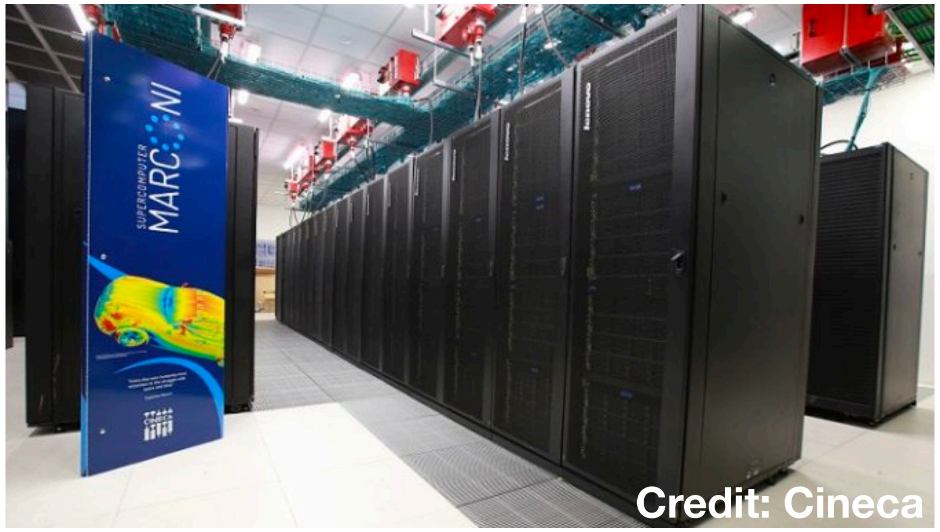
G. Bianconi
(QMU London)

M. Oberthaler's
group
(U. Heidelberg)

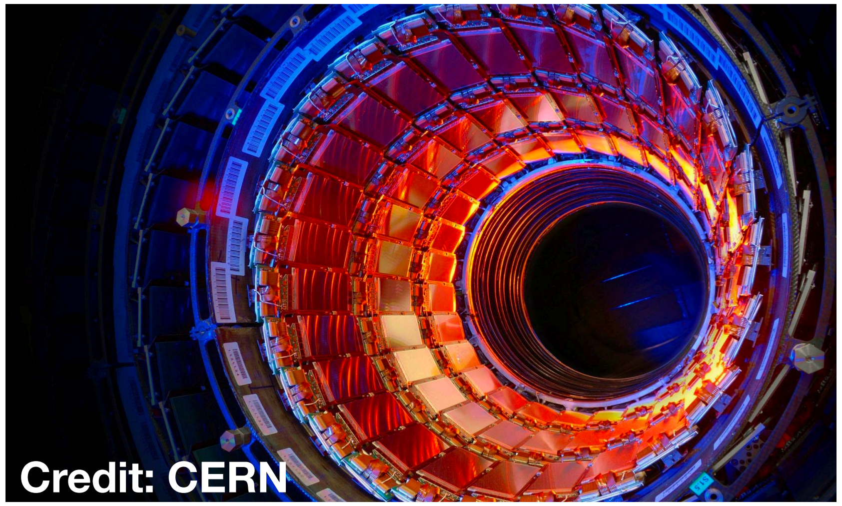
Physics in the age of data science

- Astrophysical observations

- Large-scale classical simulations



- Particle physics experiments



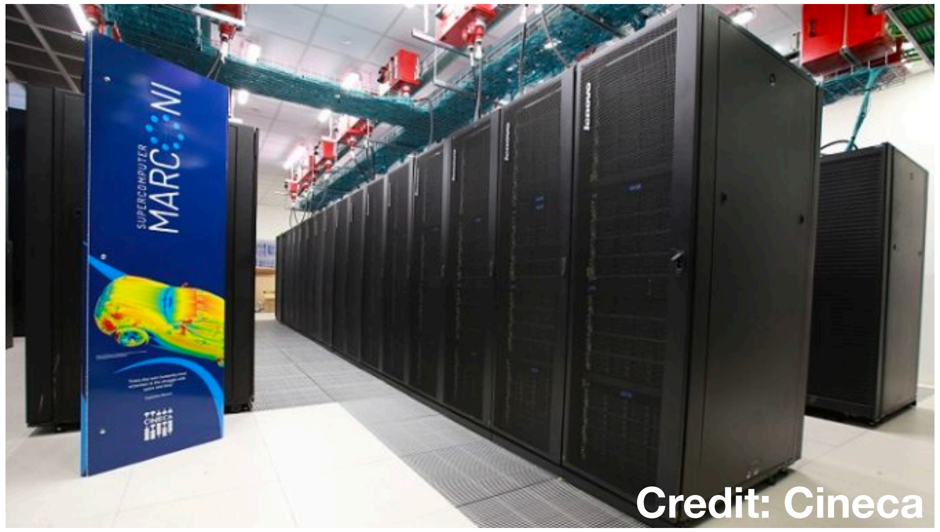
- Quantum simulation

Physics in the age of data science

<u>In this talk:</u> I'll consider classical/quantum simulation output of stat mech/many-body problems

Nowadays, these approaches can grant us access to large volumes of "many-body snapshots" (though, $N_{snapshots} \ll 2^N$)

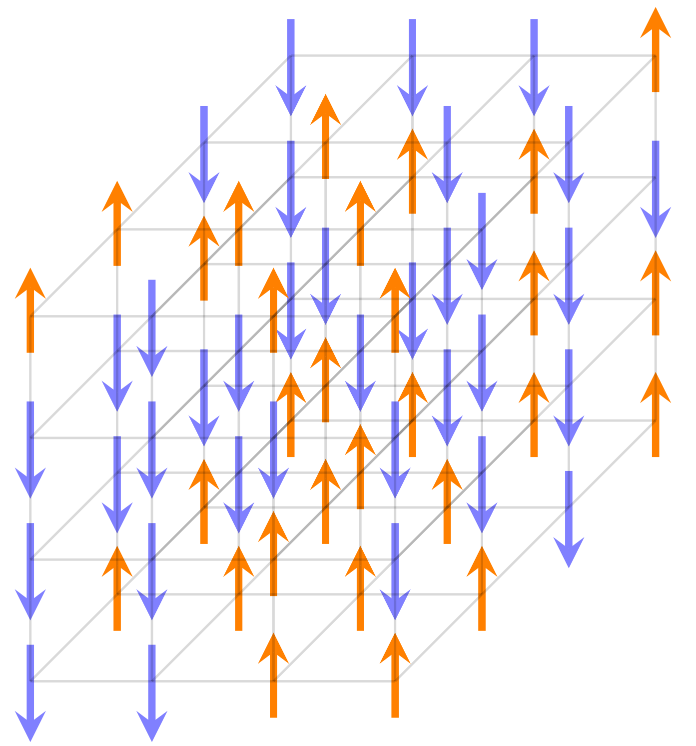
- Large-scale classical simulations



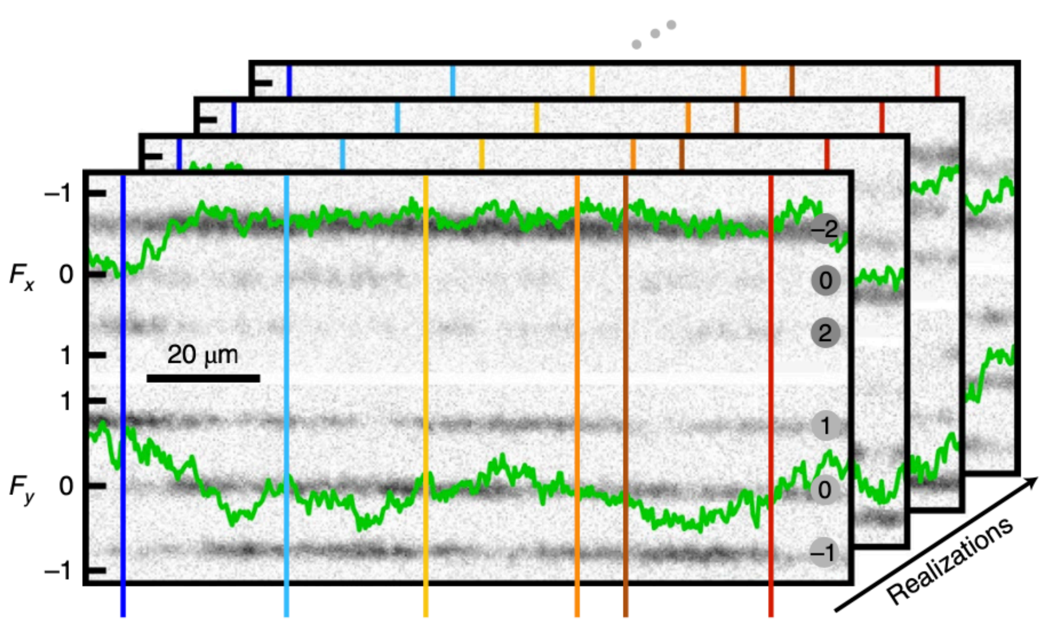
- Quantum simulation

What are many-body snapshots?

Example 1 (Stat Mech) Thermal and uncorrelated raw spin configurations sampled via Monte Carlo



Example 2 (Quantum simulation) Generalized projective measurements in a quantum simulator (e.g. local occupations)



Prüfer et al., Nat. Phys. '20

How do we extract relevant information from many-body snapshots?

"Traditional" approaches (stat mech / effective field theory): compute few-point correlators, for instance:

$$C_{ij}^{(2)} = \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle$$

Allows us to characterize classical/quantum phase transitions, determine "proper vertices" of the quantum effective action, etc.

However, it disregards part of the information content of many-body snapshots

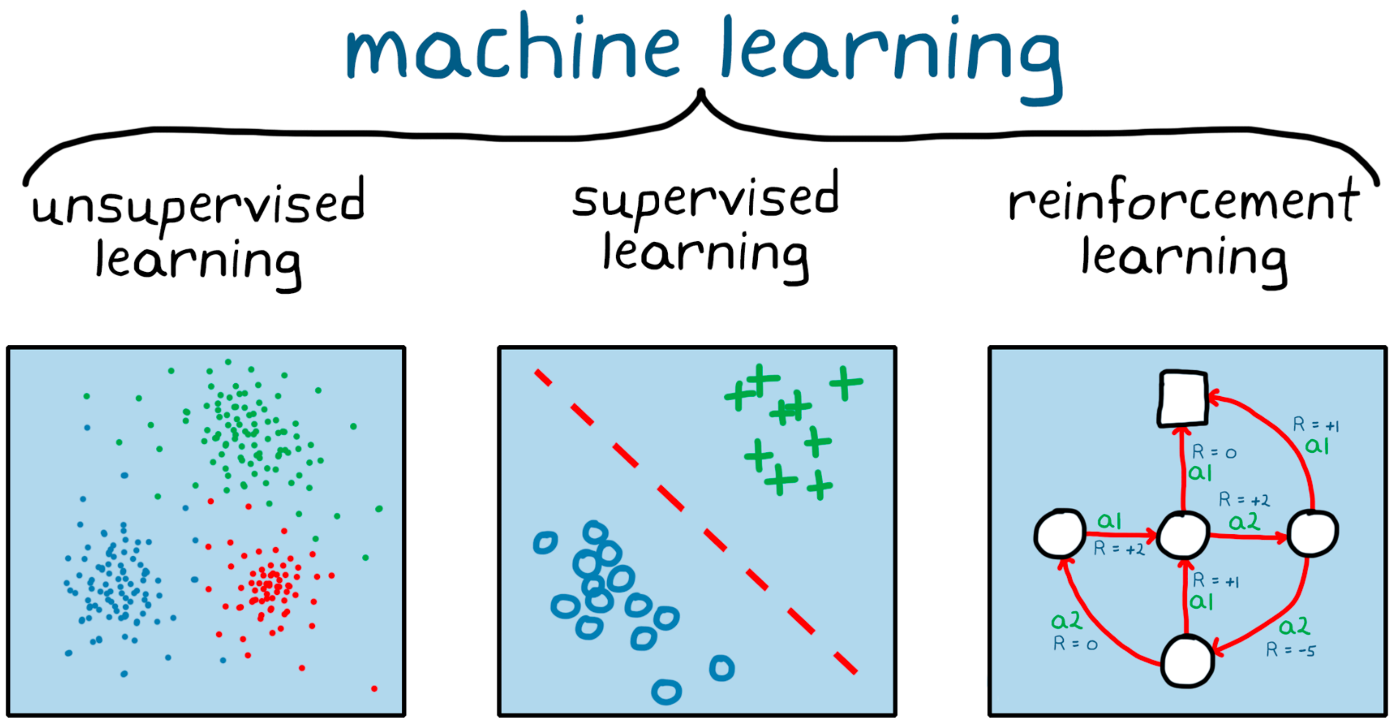
In data science jargon: an "uncontrolled" dimensional reduction

Why we would like to go beyond?

- Lattice gauge theory and topological phases (non-local correlations)
- Identifying the relevant degrees of freedom at play
- Understanding the working of quantum computers (e.g. choosing best suited observations, cross-platform verification, noise tomography, etc.)
- Quantifying the complexity of wave functions

Data-driven strategy

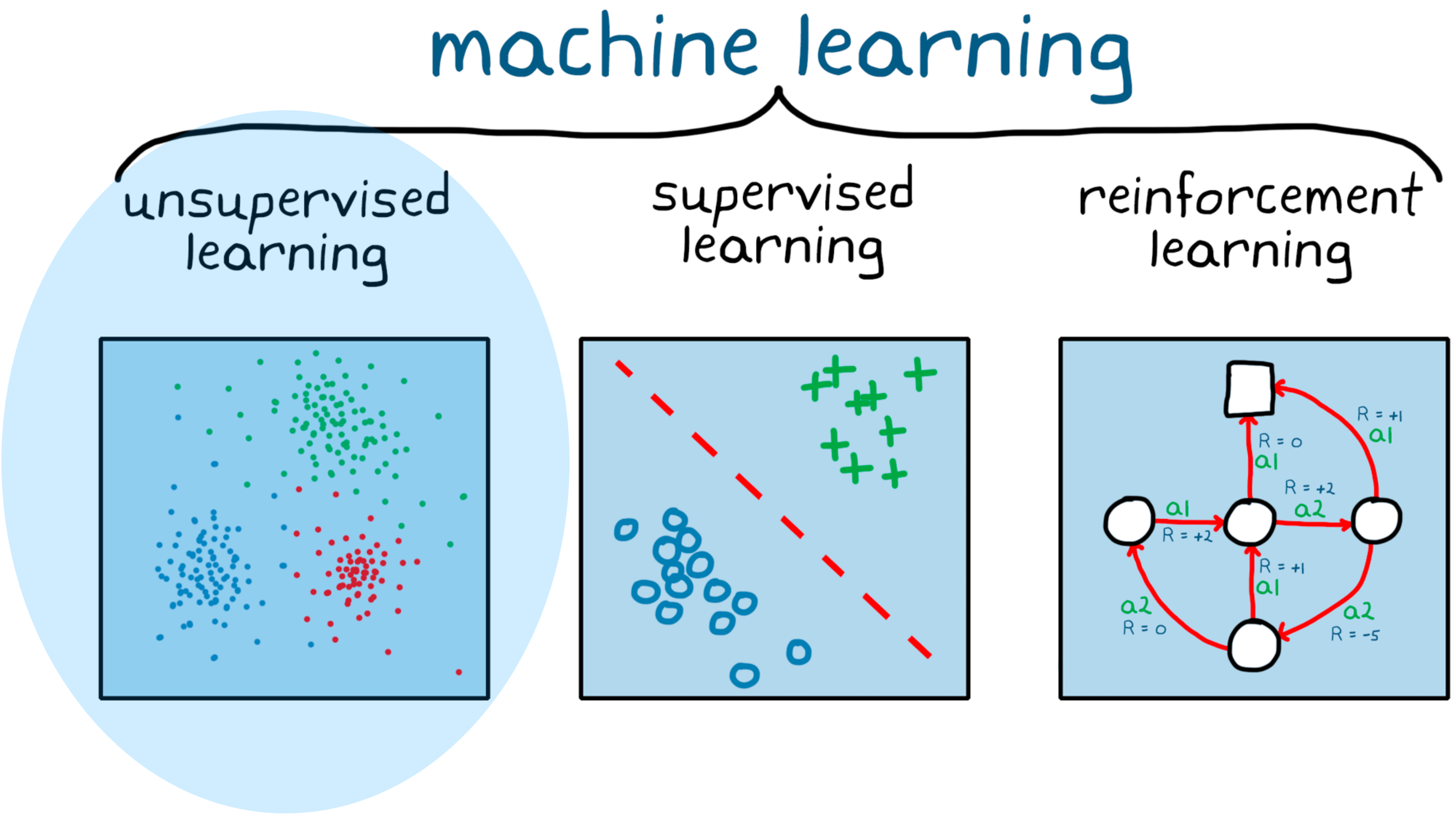
Use non-parametric statistical learning (unsupervised ML) to discover and extract relevant information in many-body physics problems by leveraging all available information



https://www.mathworks.com/discovery/reinforcement-learning.html

Data-driven strategy

Use non-parametric statistical learning (unsupervised ML) to discover and extract relevant information in many-body physics problems by leveraging all available information



https://www.mathworks.com/discovery/reinforcement-learning.html

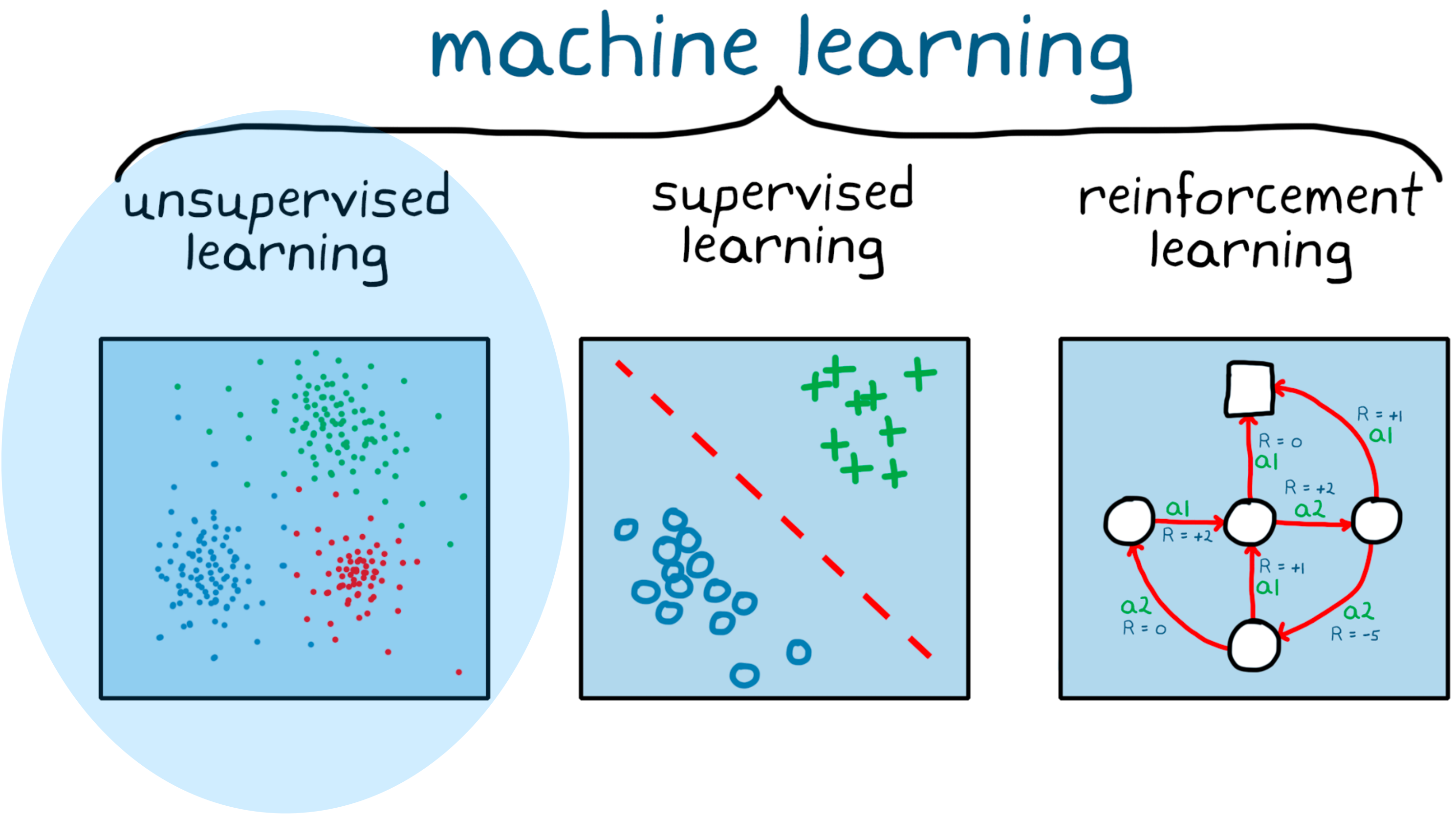
Data-driven strategy

Use non-parametric statistical learning (unsupervised ML) to discover and extract relevant information in many-body physics problems by leveraging all available information

Today, I will focus on two tools:

1) Intrinsic dimension

2) PCA entropy



https://www.mathworks.com/discovery/reinforcement-learning.html

Similar approaches

Stat mech / lattice field theory:

Hu et al. PRE '17
Wetzel PRE '17
Wang & Zhai, PRB '17
Ch'ng et al., PRE '18
Mendes-Santos et al., PRX '21
Sale et al., PRE '22; PRD '23
Sehayek & Melko, PRB '22
Spitz, et al., PRD '23
Vitale et al., arXiv '23

Quantum many-body:

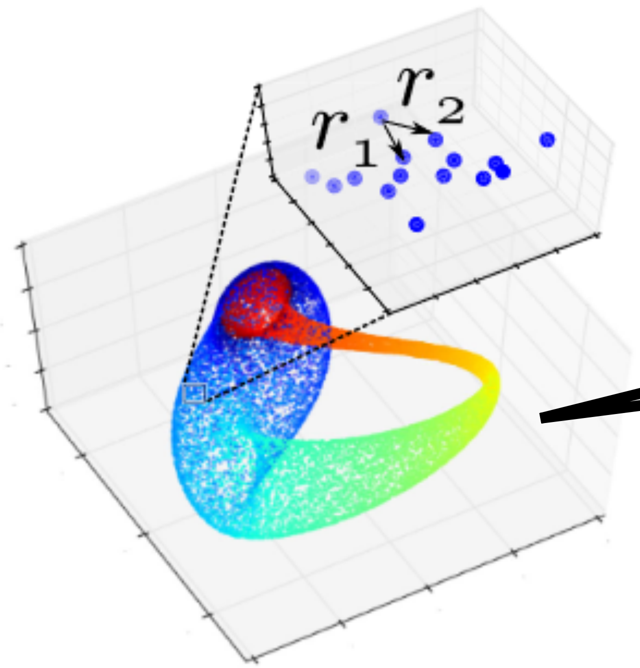
Rodriguez-Nieva & Scheurer, Nat Phys '19
Lidiak & Gong, PRL '20
Mendes-Santos et al., PRX Quantum '21
Bohrdt et al., PRL '21
Spitz, et al., SciPost Phys '21
Tirelli & Costa, PRB '21
Schmitt & Lenarčič, PRB '22
Miles et al., PRR '23
Mendes-Santos et al., arXiv '23

… and many more!

Intrinsic dimension

- Basic tool in data mining with multiple applications in chemical and bimolecular science and image analysis
- Quantifies the minimum number of variables needed to describe the data
- Serves as a proxy of the Kolmogorov complexity

$D = 3$



$I_d = 2$

Each point is, for example, a spin configuration

Mendes-Santos et al., PRX ‘21

Intrinsic dimension: TWO-NN

Facco et al., Sci. Rep. '17

Uses statistics of distances between nearest-neighbor (NN) points

Needs a metric (e.g. for spin systems: Hamming distance)

$$d(i,j) := \sum_r |\vec{S}^i_r - \vec{S}^j_r|$$

Intrinsic dimension: TWO-NN

Facco et al., Sci. Rep. '17

Uses statistics of distances between nearest-neighbor (NN) points

Needs a metric (e.g. for spin systems: Hamming distance)

$$d(i,j) := \sum_r |\vec{S}^i_r - \vec{S}^j_r|$$

Example: 3-site system

$$\vec{S}^1 = (0,1,1) \qquad d(\vec{S}^1, \vec{S}^2) = |0-1| + |1-1| + |1-1| = 1$$

$$\vec{S}^2 = (1,1,1) \qquad d(\vec{S}^1, \vec{S}^3) = 2$$

$$\vec{S}^3 = (1,0,1) \qquad \ldots$$

$$\vec{S}^4 = (0,0,0)$$

Intrinsic dimension: TWO-NN

Facco et al., Sci. Rep. '17

Uses statistics of distances between nearest-neighbor (NN) points

Needs a metric (e.g. for spin systems: Hamming distance)

Main assumption: NN points are drawn uniformly from I_d-dim hyperspheres

For each point, compute:

$$\mu = \frac{r_2}{r_1}$$

Distribution function of μ:

$$f(\mu) = \frac{I_d}{\mu^{I_d+1}}$$

Intrinsic dimension: TWO-NN

Facco et al., Sci. Rep. '17

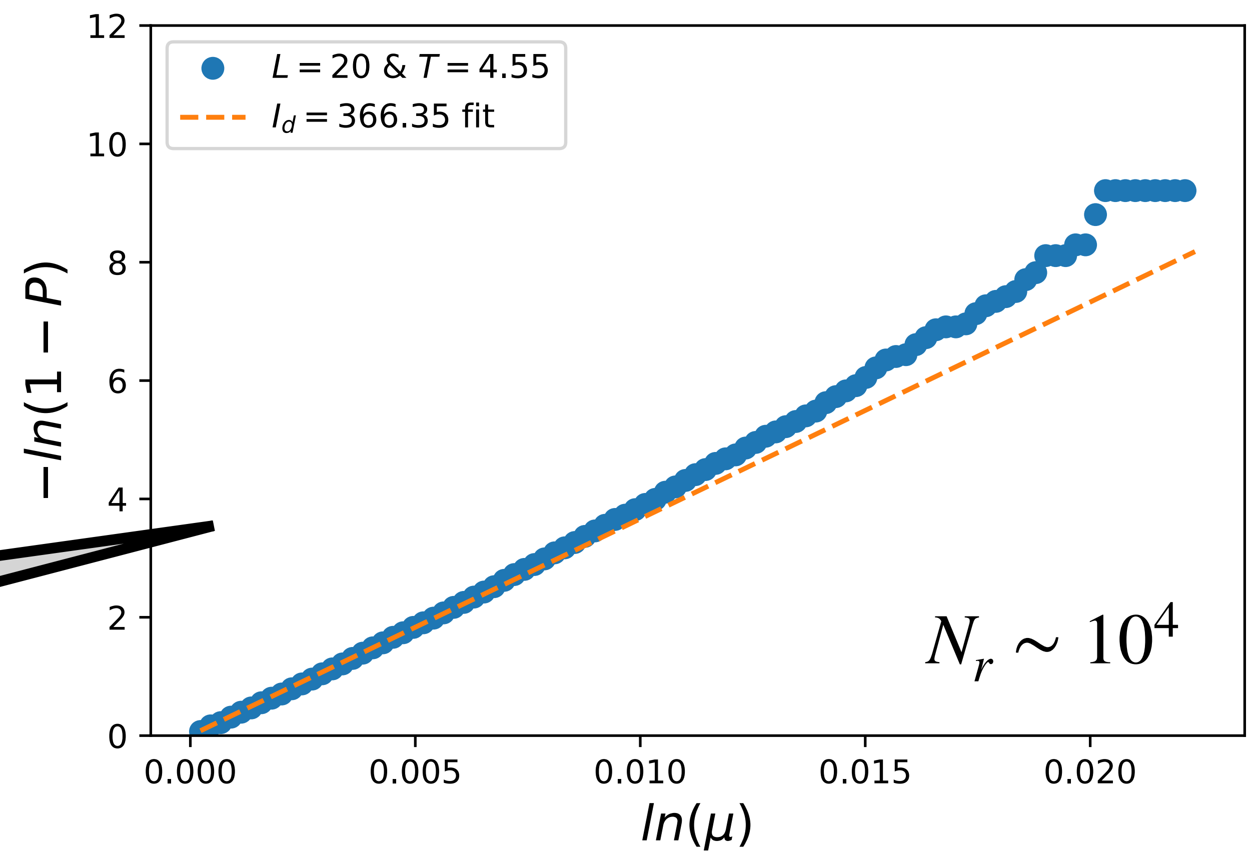
Uses statistics of distances between nearest-neighbor (NN) points

Needs a metric (e.g. for spin systems: Hamming distance)

Main assumption: NN points are drawn uniformly from I_d-dim hyperspheres

$$\mu = \frac{r_2}{r_1} \qquad f(\mu) = \frac{I_d}{\mu^{I_d+1}}$$

Linear fit using cumulative dist. function



Intrinsic dimension: toy example

<u>Toy example: 3-site XY model</u>

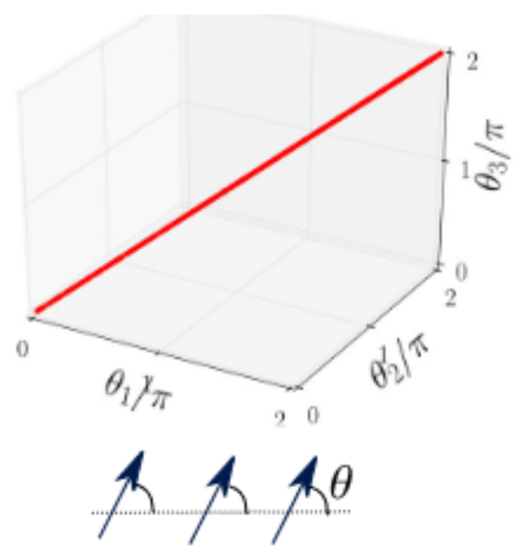
Mendes-Santos et al., PRX '21

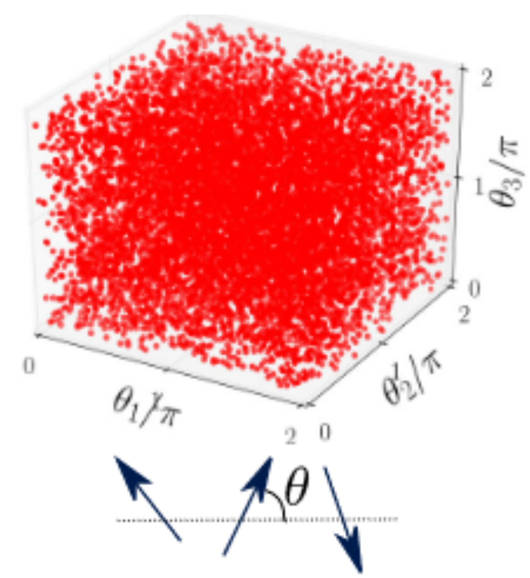
Hamiltonian: $$H = -\sum_{\langle i,j \rangle} \cos(\theta_i - \theta_j)$$

Configurations (data points): $(\theta_1, \theta_2, \theta_3)$

Low temperature

$$I_d = 1$$





High temperature

$$I_d = D = 3$$

What about close to a transition point?

Intrinsic dimension: 2D Ising

2D classical Ising model

$$E = -J \sum_{\langle i,j \rangle} S_i S_j$$

Square lattice

Divergent correlation length: do data structures are more complex?

Second-order (conformal) phase transition

$$T_c = \frac{2}{\ln(1+\sqrt{2})} \approx 2.269$$

$$\nu = 1$$

Mendes-Santos et al., PRX '21

Intrinsic dimension: 2D Ising

<u>2D classical Ising model</u>

$$E = -J \sum_{\langle i,j \rangle} S_i S_j$$

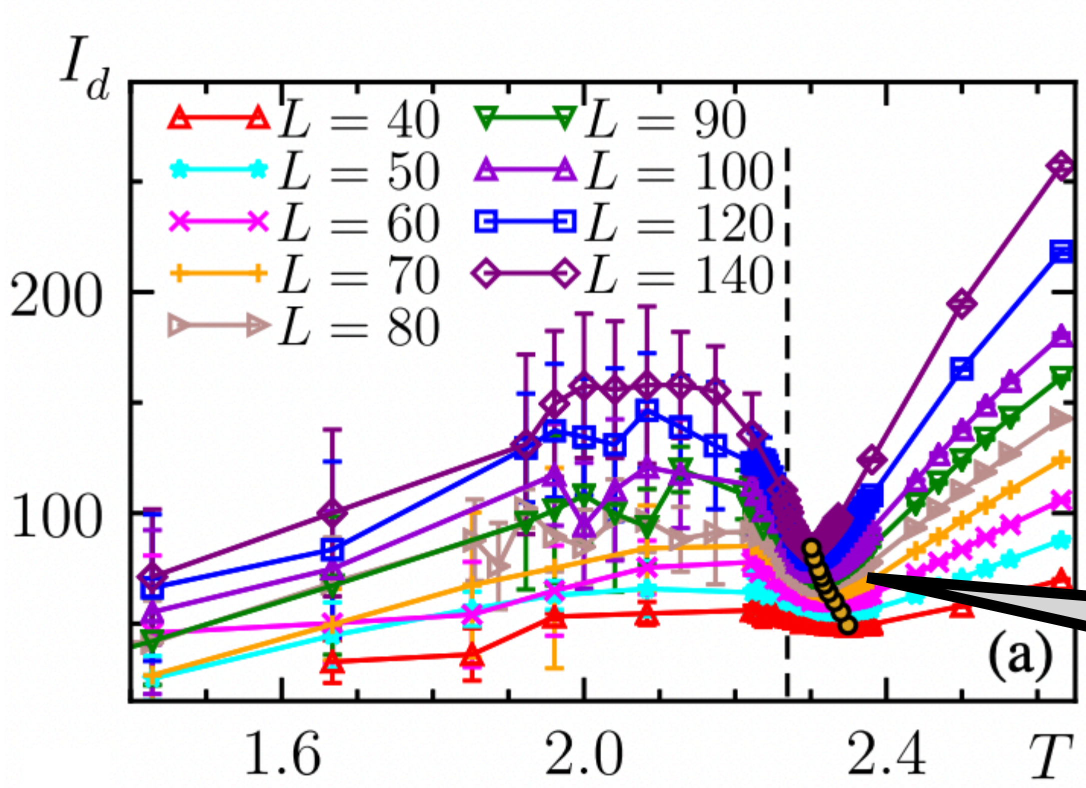
Square lattice

Second-order (conformal) phase transition

$$T_c = \frac{2}{\ln(1+\sqrt{2})} \approx 2.269$$

$$\nu = 1$$

Divergent correlation length: do data structures are more complex?



Manifold simplifies at the transition!

Intuition: universality

Mendes-Santos et al., PRX '21

Intrinsic dimension: 2D Ising

<u>2D classical Ising model</u>

$$E = -J \sum_{\langle i,j \rangle} S_i S_j$$

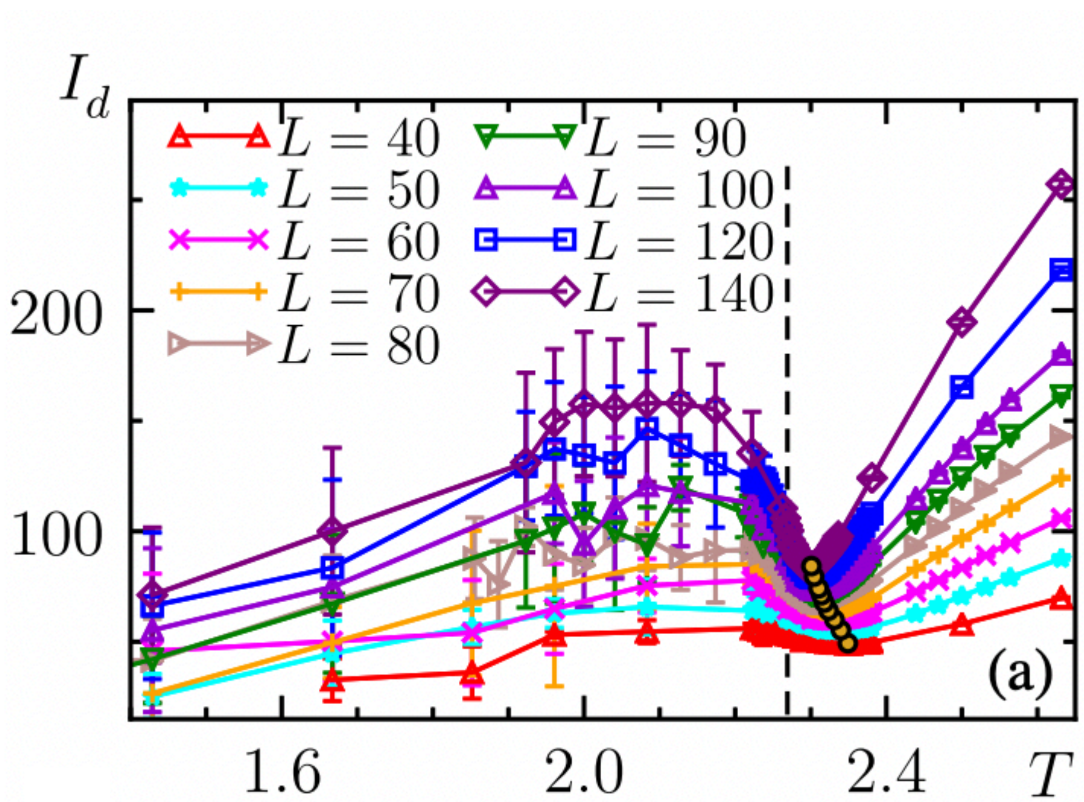
Square lattice

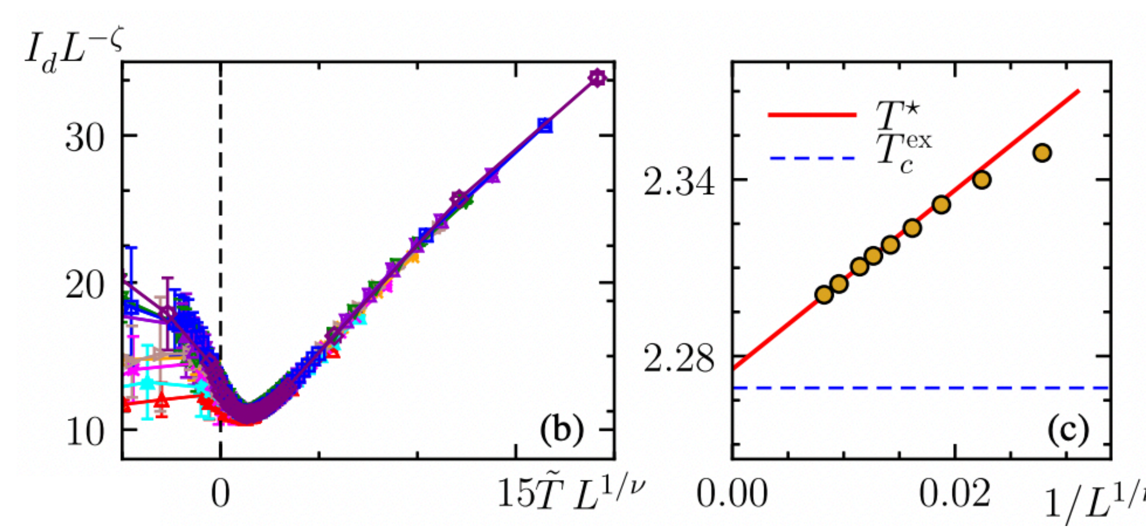
Divergent correlation length: do data structures are more complex?

Second-order (conformal) phase transition

$$T_c = \frac{2}{\ln(1+\sqrt{2})} \approx 2.269$$

$$\nu = 1$$





Mendes-Santos et al., PRX '21

Role of the physical dimension

How does the physical dimension affects the data structure and the intrinsic dimension?

Rajat Panda

3D Ising model $$E = -J \sum_{\langle i,j \rangle} S_i S_j$$

- No analytical solution known so far
- Continuous phase transition at $T_c \approx 4.51$ (believed to be conformal)
- Dual to a $\mathbb{Z}_2$ lattice gauge theory
- QCD critical point expected to belong to the 3D Ising universality class
  [Stephanov et al., PRL '98; Gavin et al., PRD '94; ...]

Panda, RV, et al., in preparation

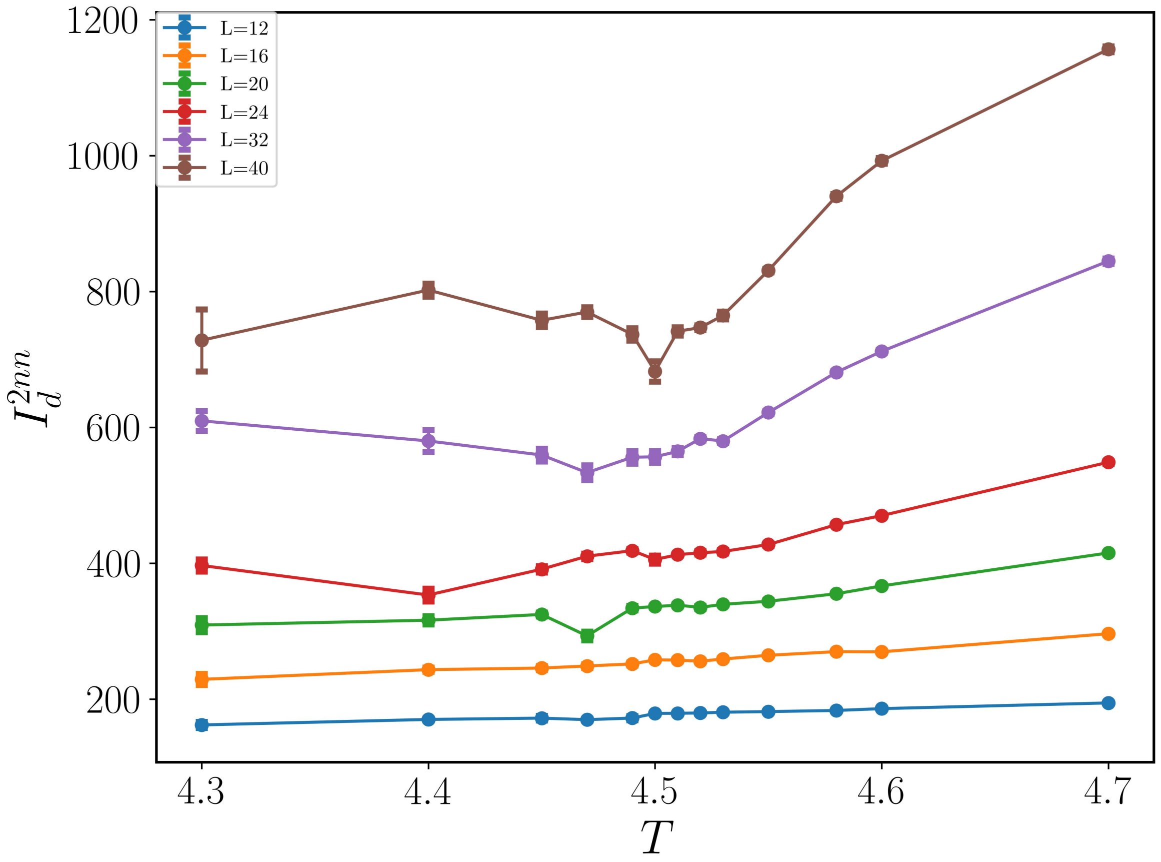
Role of the physical dimension

How does the physical dimension affects the data structure and its intrinsic dimension?

Rajat Panda

3D Ising model

$$E = -J \sum_{\langle i,j \rangle} S_i S_j$$



- Very high I_d (results must be taken warily)
- Minimum not so clear at the transition (TWO-NN estimator)
- In general, harder to extract information through I_d

Panda, RV, et al., in preparation

PCA entropy

Can we use complementary statistical tests to still be able to extract relevant information?

PCA entropy

Can we use complementary statistical tests to still be able to extract relevant information?

Principal Component Analysis (PCA)

Transformation of the coordinate system to find high-variance directions

It amounts to diagonalizing the covariance matrix $\mathbf{\Sigma} = \mathbf{X}^T\mathbf{X}/(N_r - 1)$:

$$\mathbf{\Sigma}\lambda_n = \lambda_n \overrightarrow{w}_n$$

See e.g. Jolliffe (2005)

PCA entropy

Can we use complementary statistical tests to still be able to extract relevant information?

Principal Component Analysis (PCA)

Transformation of the coordinate system to find high-variance directions

It amounts to diagonalizing the covariance matrix $\mathbf{\Sigma} = \mathbf{X}^T\mathbf{X}/(N_r - 1)$:

$$\mathbf{\Sigma}\lambda_n = \lambda_n \overrightarrow{w}_n$$

See e.g. Jolliffe (2005)

Normalized eigenvalues:

$$\tilde{\lambda}_n = \frac{\lambda_n}{\sum_m \lambda_m}$$

By construction: $\tilde{\lambda}_n \geq 0, \quad \sum_n \tilde{\lambda}_n = 1$

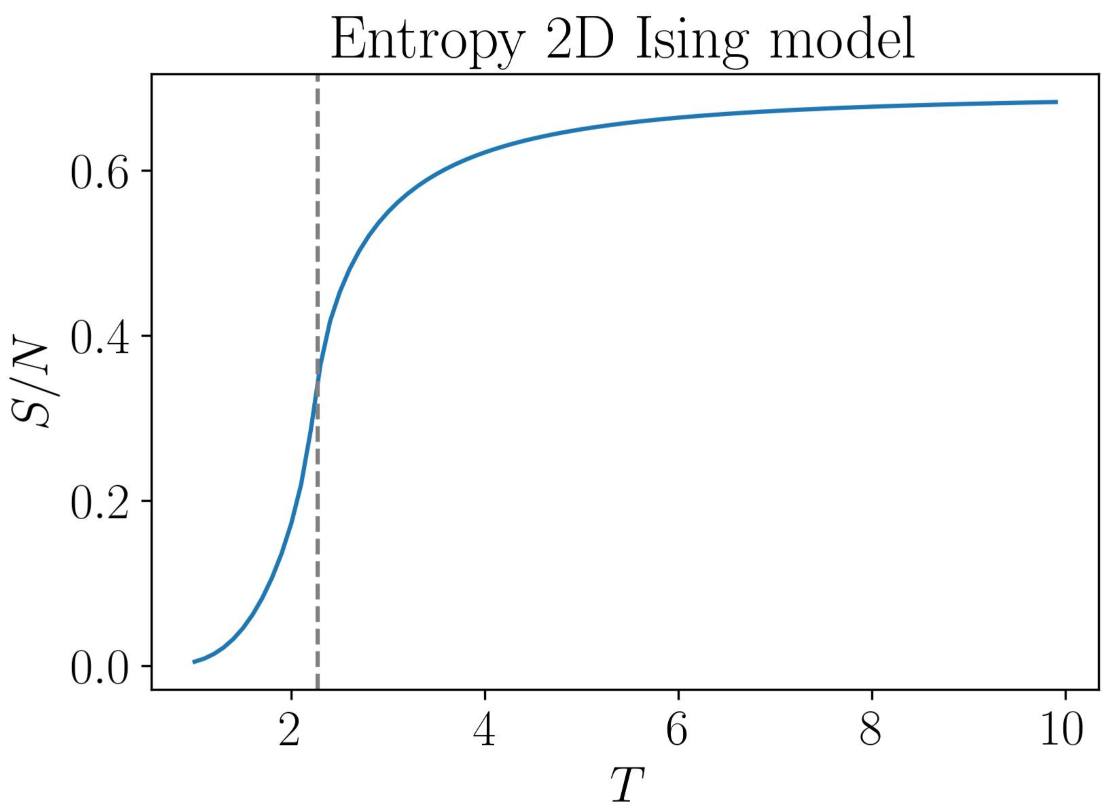
("Shannon") PCA entropy

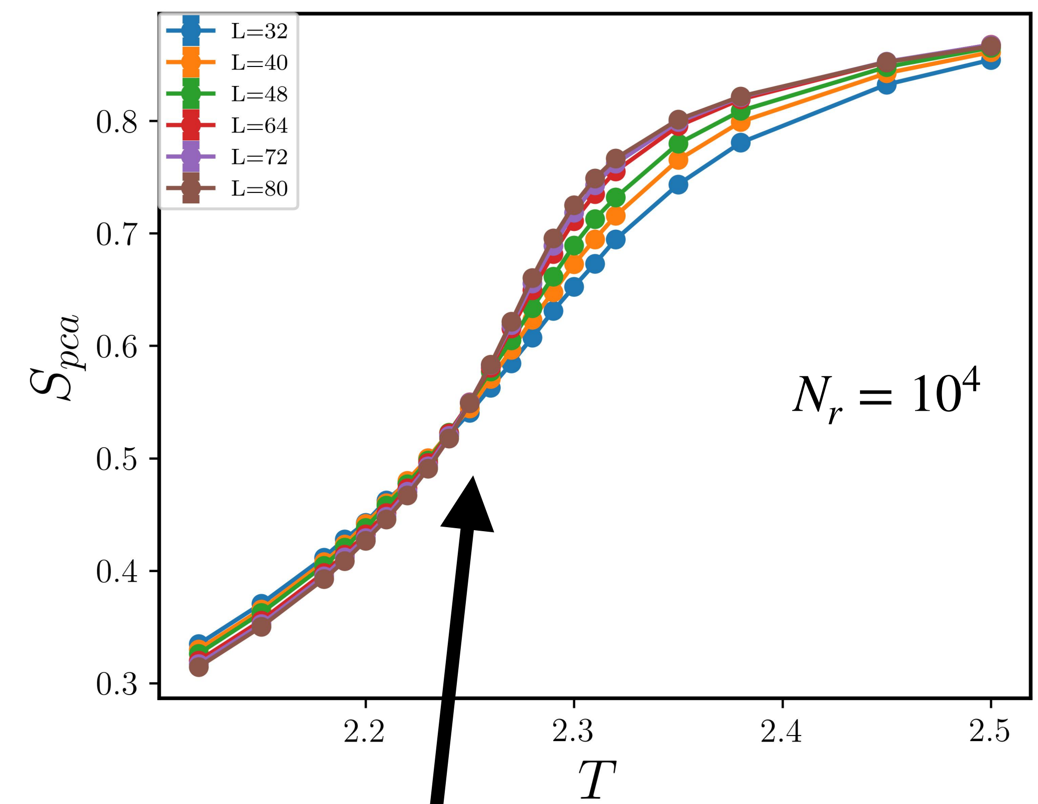
$$S_{\mathrm{PCA}} = -\sum_n \tilde{\lambda}_n \ln(\tilde{\lambda}_n)$$

Alter et al., PNAS (2000), …

PCA entropy: 2D Ising

Striking qualitative similarity to the thermodynamic entropy!



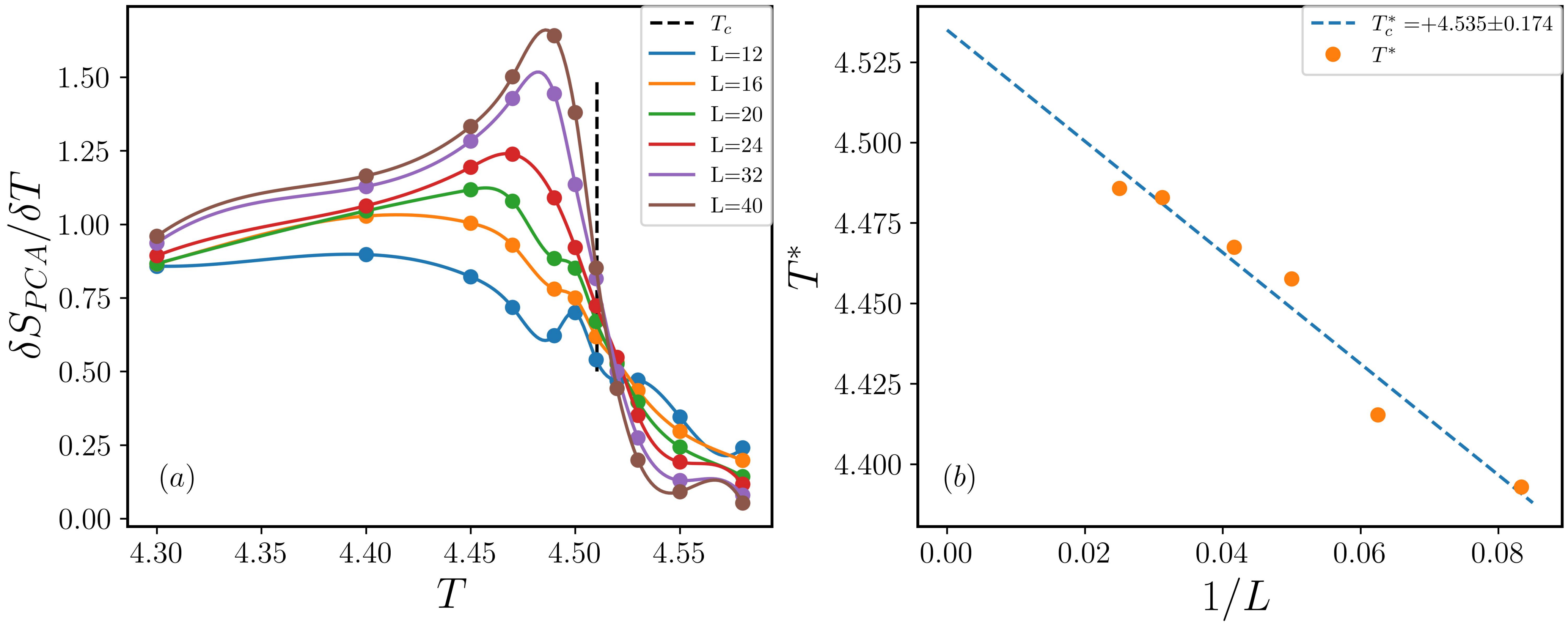


Flex very close to the transition point

Panda, RV, et al., in preparation

PCA entropy: 3D Ising

Also works nicely for the 3D model!



Allows to estimate T_c with less than 1% error

Panda, RV, et al., in preparation

Experiments

LETTERS

https://doi.org/10.1038/s41567-020-0933-6

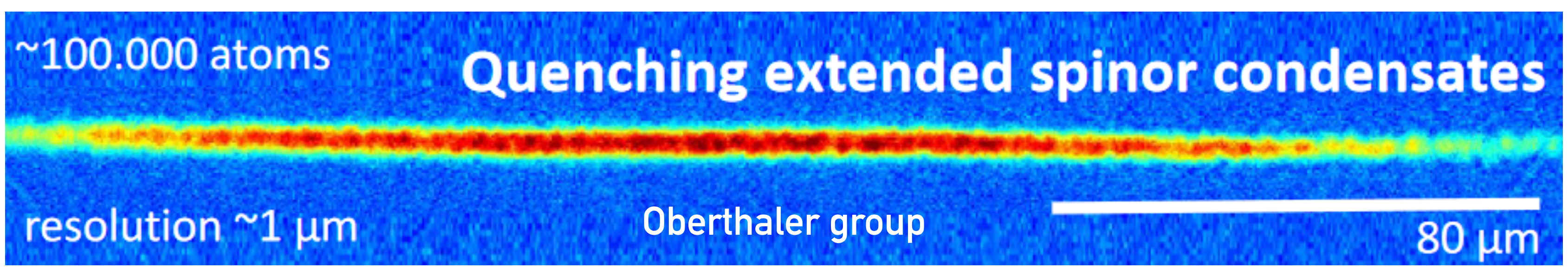
nature physics

Check for updates

Experimental extraction of the quantum effective action for a non-equilibrium many-body system

Maximilian Prüfer[1,3], Torsten V. Zache[2,3], Philipp Kunkel[1], Stefan Lannig[1], Alexis Bonnin[1], Helmut Strobel[1], Jürgen Berges[2] and Markus K. Oberthaler[1]

In collaboration with M. Oberthaler's group



$$\Gamma_t[\Phi] = \sum_{n=1}^{\infty} \frac{1}{n!} \Gamma_t^{\alpha_1, \dots, \alpha_n}(y_1, \dots, y_n)\, \Phi^{\alpha_1}(y_1) \cdots \Phi^{\alpha_n}(y_n)$$

$$\downarrow$$

What are the relevant operators to determine the proper vertices?

Experiments

LETTERS

https://doi.org/10.1038/s41567-020-0933-6

nature physics

Check for updates

Experimental extraction of the quantum effective action for a non-equilibrium many-body system

Maximilian Prüfer[1,3] ✉, Torsten V. Zache[2,3], Philipp Kunkel[1], Stefan Lannig[1], Alexis Bonnin[1], Helmut Strobel[1], Jürgen Berges[2] and Markus K. Oberthaler[1]

In collaboration with M. Oberthaler's group

$$\Gamma_t[\Phi] = \sum_{n=1}^{\infty} \frac{1}{n!} \Gamma_t^{\alpha_1,\dots,\alpha_n}(y_1,\dots,y_n)\, \Phi^{\alpha_1}(y_1)\cdots\Phi^{\alpha_n}(y_n)$$

Obtained from irreducible parts of correlators of the transverse spin

$$F_\perp(y) = F_x(y) + iF_y(y) = |F_\perp(y)|\,e^{i\varphi(y)}$$

See e.g. Kawaguchi & Ueda, Phys. Rep. '12

Experiments

LETTERS

https://doi.org/10.1038/s41567-020-0933-6

nature physics

Check for updates

Experimental extraction of the quantum effective action for a non-equilibrium many-body system

Maximilian Prüfer[1,3], Torsten V. Zache[2,3], Philipp Kunkel[1], Stefan Lannig[1], Alexis Bonnin[1], Helmut Strobel[1], Jürgen Berges[2] and Markus K. Oberthaler[1]

In collaboration with M. Oberthaler's group

$$\Gamma_t[\Phi] = \sum_{n=1}^{\infty} \frac{1}{n!} \Gamma_t^{\alpha_1,\dots,\alpha_n}(y_1,\dots,y_n)\, \Phi^{\alpha_1}(y_1)\cdots\Phi^{\alpha_n}(y_n)$$

Obtained from irreducible parts of correlators of the transverse spin

Determined by particular combinations of populations

$$F_\perp(y) = F_x(y) + iF_y(y) = |F_\perp(y)|\mathrm{e}^{i\varphi(y)}$$

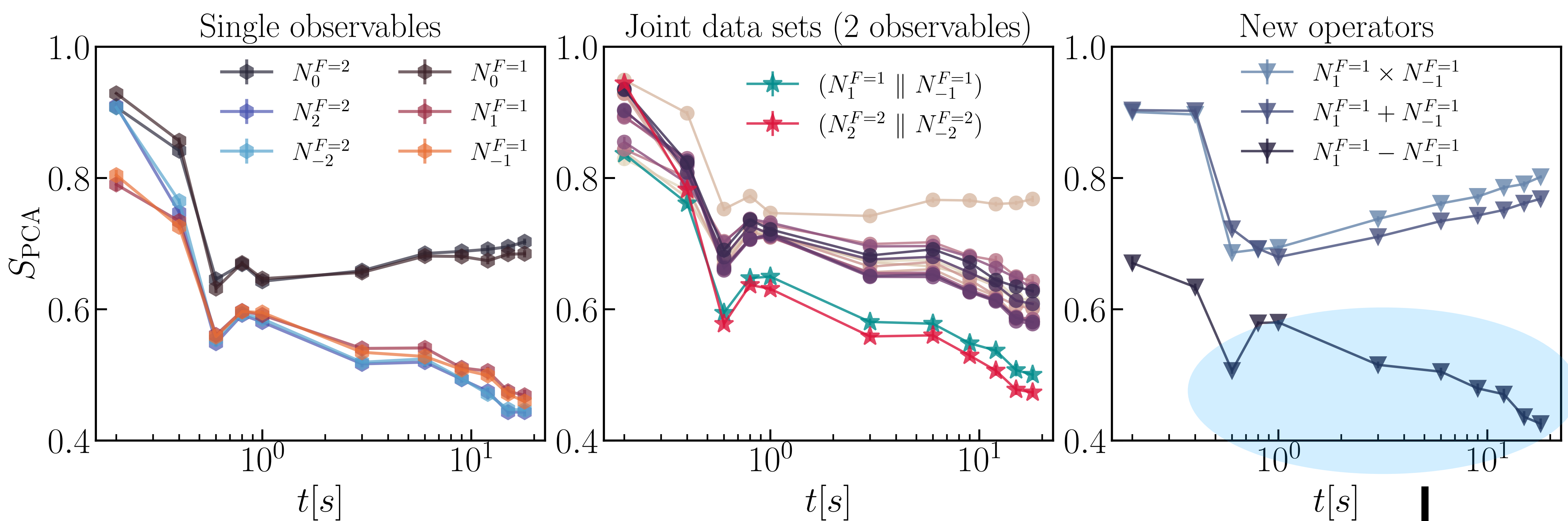
$$F_x(y) = \left(N_{+2}^{F=2}(y) - N_{-2}^{F=2}(y)\right)/N_{\mathrm{tot}}^{F=2}(y)$$

$$F_y(y) = \left(N_{+1}^{F=1}(y) - N_{-1}^{F=1}(y)\right)/N_{\mathrm{tot}}^{F=1}(y)$$

See e.g. Kawaguchi & Ueda, Phys. Rep. '12

Ranking of observables

PCA entropy provides a metric to rank observables based on their relevance: the lower S_{PCA} the stronger the correlations within an observation



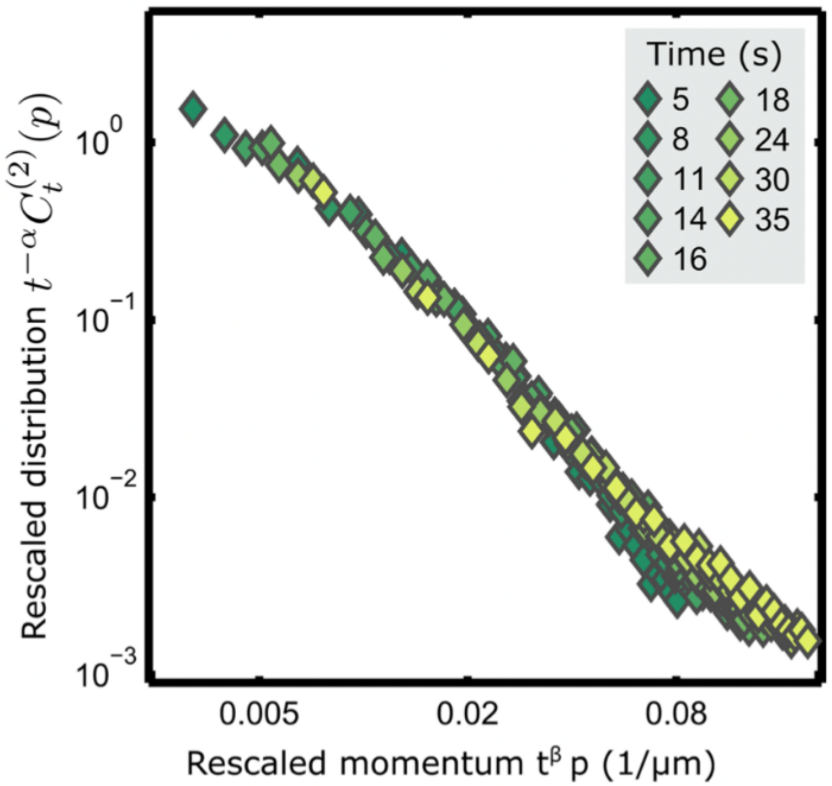
$$F_x(y) = \left(N_{+2}^{F=2}(y) - N_{-2}^{F=2}(y)\right)/N_{\mathrm{tot}}^{F=2}(y)$$

$$F_y(y) = \left(N_{+1}^{F=1}(y) - N_{-1}^{F=1}(y)\right)/N_{\mathrm{tot}}^{F=1}(y)$$

RV, et al., in preparation (on arXiv soon!)

Agnostic bound on universal scaling regime

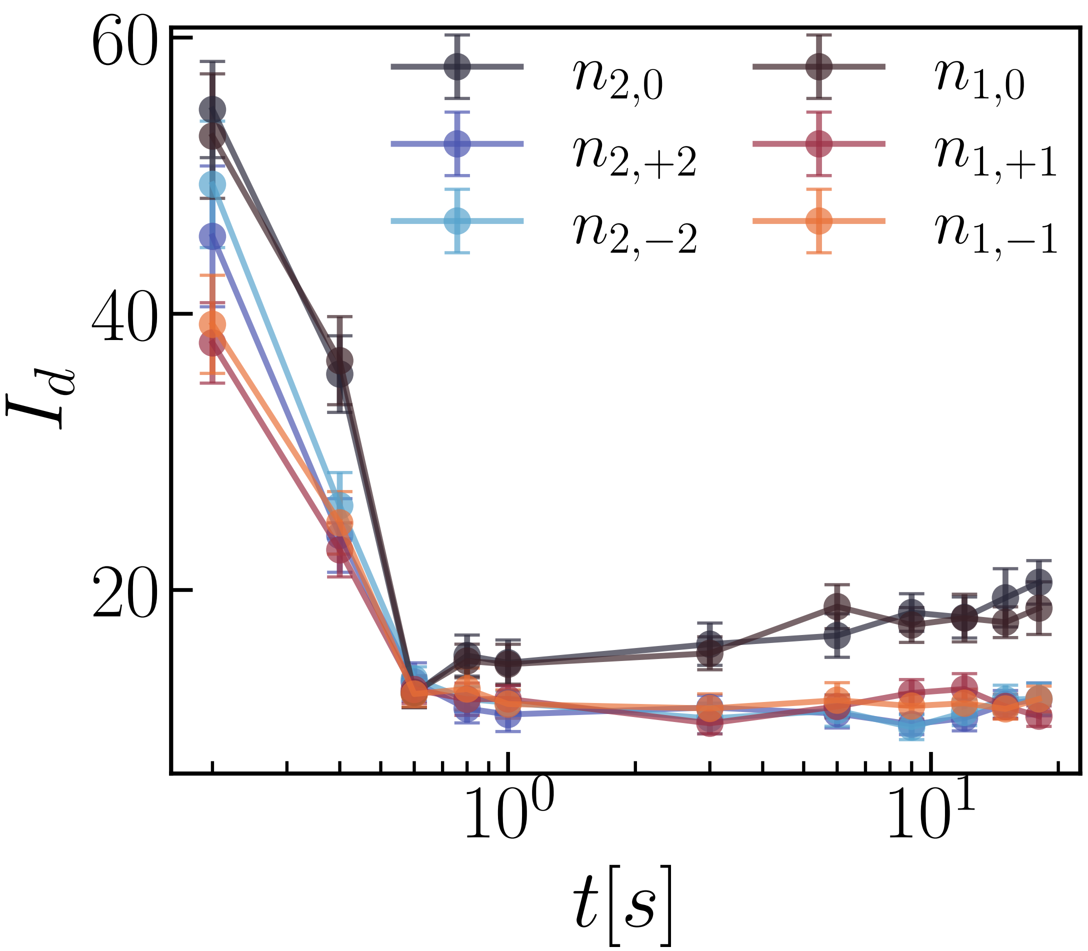
Correlation functions of the transverse spin exhibit self-similar dynamics



Prüfer et al., Nat. Phys. '20

Theo: Berges et al., PRL '08, …

Intrinsic dimension features long, stable plateaus in strong agreement with universal behavior



RV, et al., in preparation (on arXiv soon!)

Conclusions

- Non-parametric statistical learning provides powerful tools to enable assumption-free discoveries in many-body physics!
- Widely applicable methods: classical/quantum, in and out of equilibrium (working with modest volumes of data)
- Insights on lattice gauge theory and topological matter (on-going)
- Interesting connections to the entropy and measures of complexity (e.g. Kolmogorov complexity, Shannon entropy)

Thank you!

Extra material

Further applications

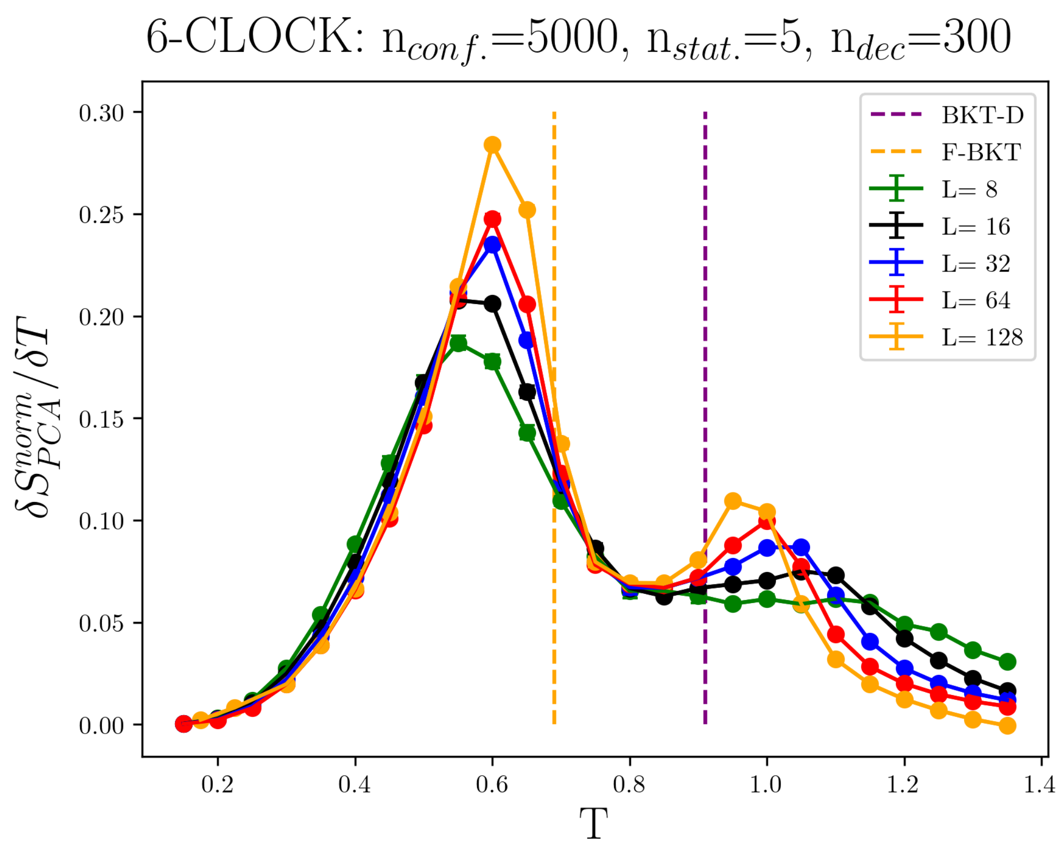
q-clock models and BKT
("discretized" XY model)

S. Pedrielli

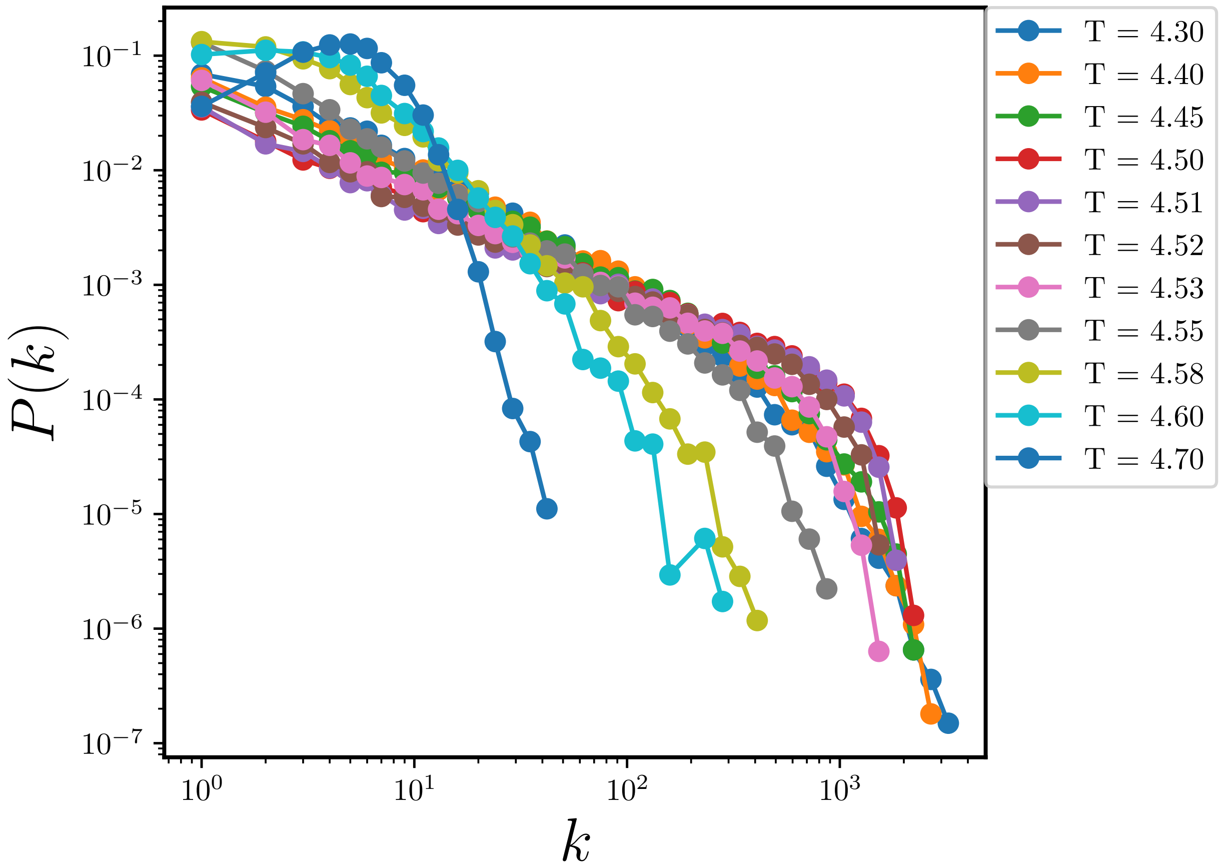
Ising partition
function networks

H. Sun

G. Bianconi



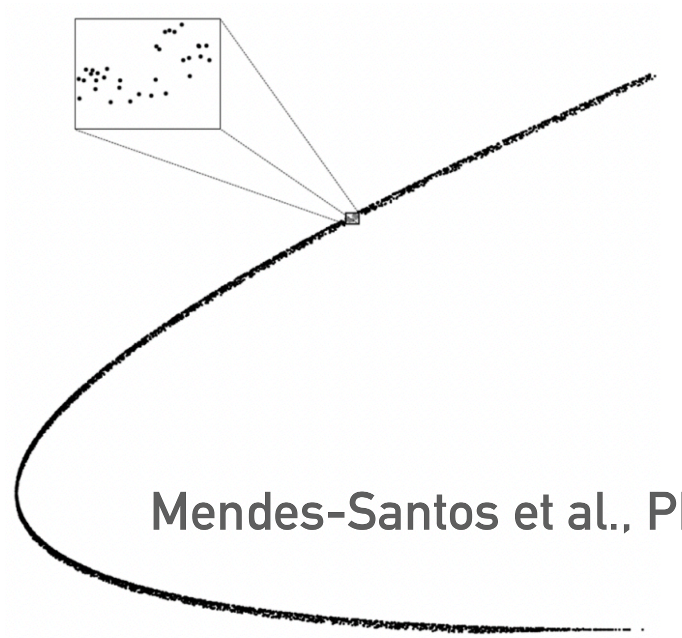
Pedrielli, RV, et al., in preparation



Sun, RV, et al., in preparation

More about intrinsic dimension

- Lower bound of complexity in data sets (e.g. relation to bottleneck in autoencoders [Ansuini et al., NearIPS 2019])
- Crucial dependence on the chosen scale



Mendes-Santos et al., PRX ‘21

- Related to the Kolmogorov complexity

How long shall a classical computer code be to reproduce a given string?

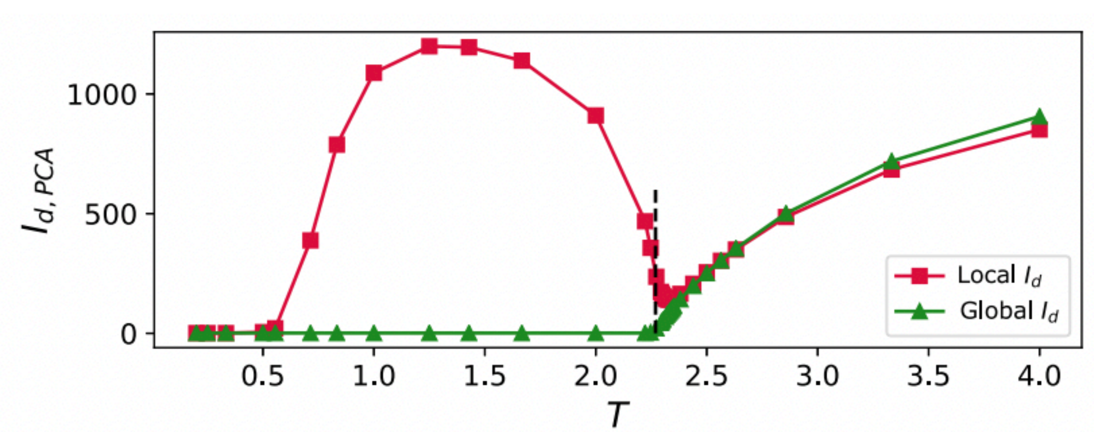
‘11111111…’	‘10011010…’
print ‘1’ n times (lower complexity)	print ‘10011010…’ (higher complexity)

I_d estimation: PCA

See e.g. Jolliffe (2005)

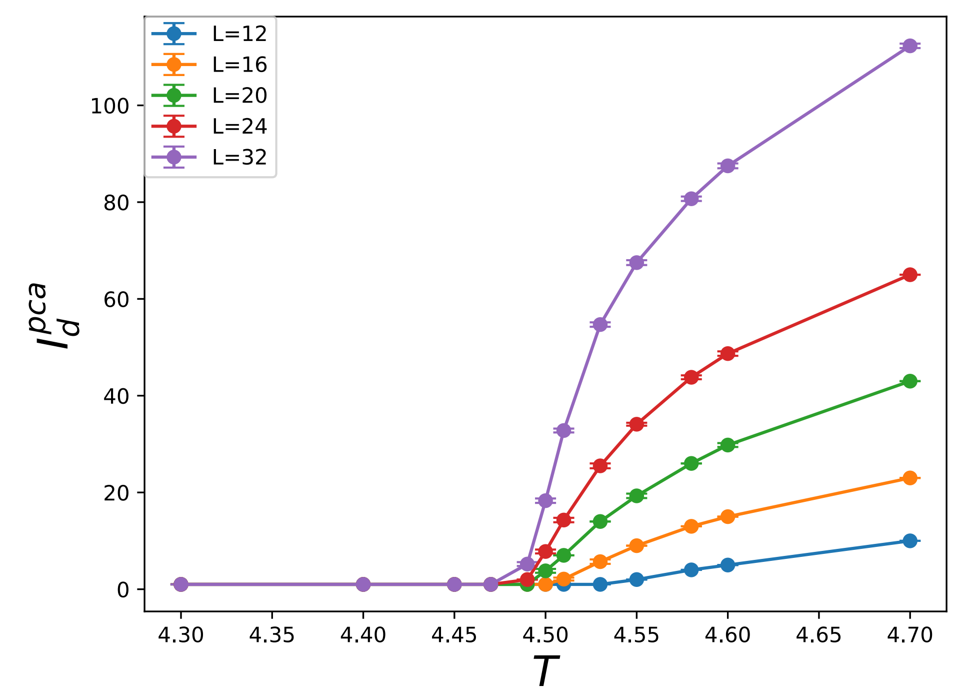
- Based on a ad-hoc cutoff parameter in the integrated spectrum of the covariance matrix $\sum_{n=1}^{I_d} \tilde{\lambda}_n \approx \zeta$
- Bad estimate for curved manifolds

2D Ising



Mendes-Santos et al., PRX '21

3D Ising

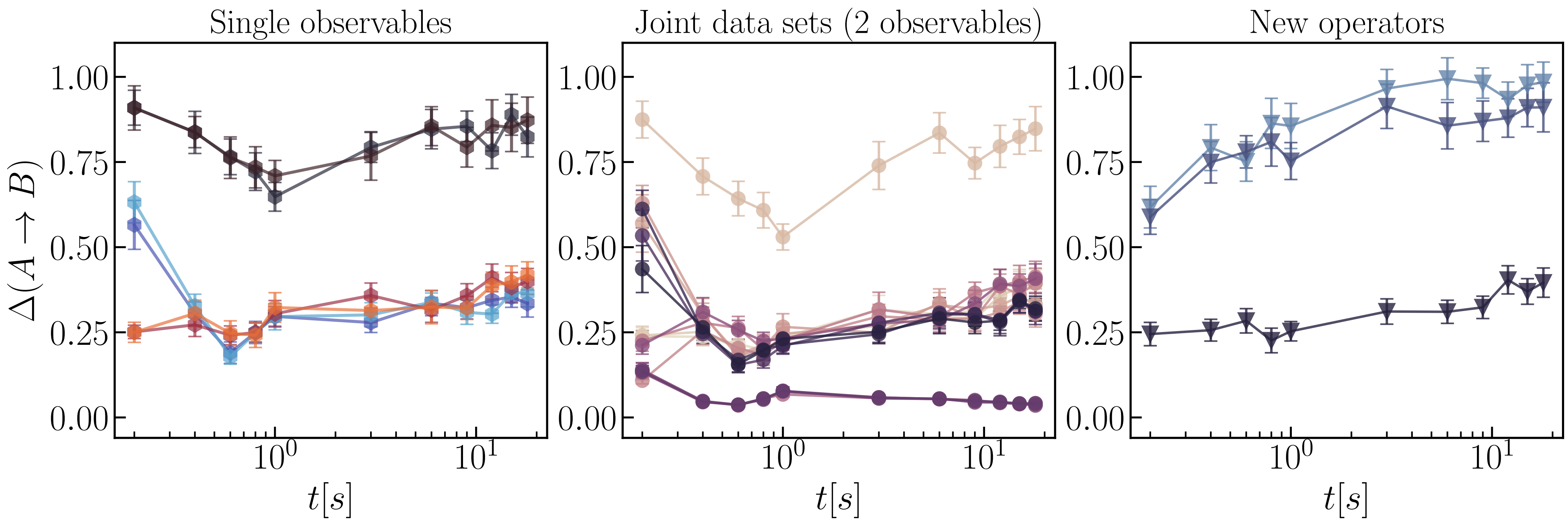


Panda, RV, et al., in preparation

Ranking of observables: information imbalance

Complementary metric of relevance used in unsupervised ML: information imbalance (based on rankings of NN distances)

Glielmo et al., PNAS Nexus '21



Fully consistent with PCA entropy prediction
(ask me later if interested in details)

RV, et al., in preparation (on arXiv soon!)