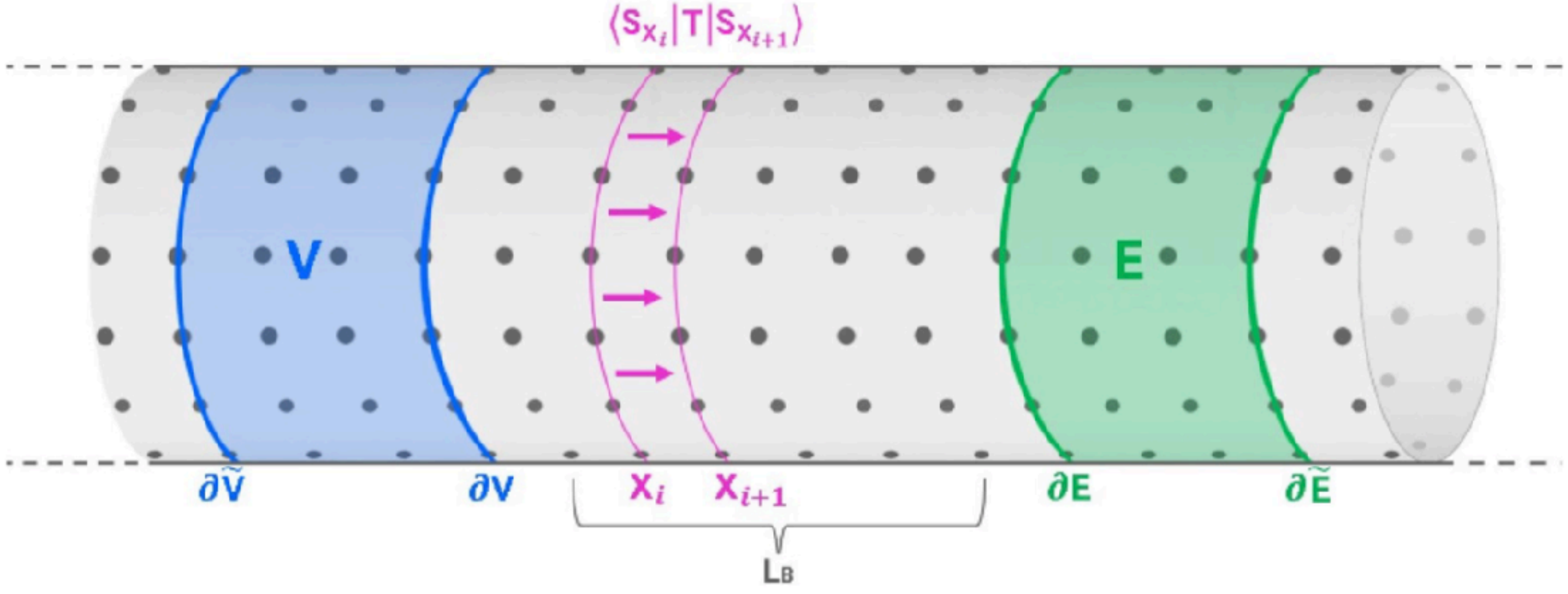
Finding slow variables using Information Bottlenecks

venturing into unsolved models



Zohar Ringel | August 2022

Maciej Koch-Janusz

Sebastian D. Huber

Patrick Lenggenhager

Doruk Efe Gokmen

University of Zurich[UZH]

ETH zürich

Maciej Koch Janusz, and ZR, *Nature Physics **14**, 578-582 (2018)*

P. Lenngenhager, D.E. Gokmen, ZR, S.D. Huber and Maciej Koch Janusz,
*Phys. Rev. X **10**, 011037 (2020)*

Amit Gordon, Aditya Banerjee, Maciej Koch Janusz and ZR,
Physical Review Letters 126 (24), 240601 (2021)

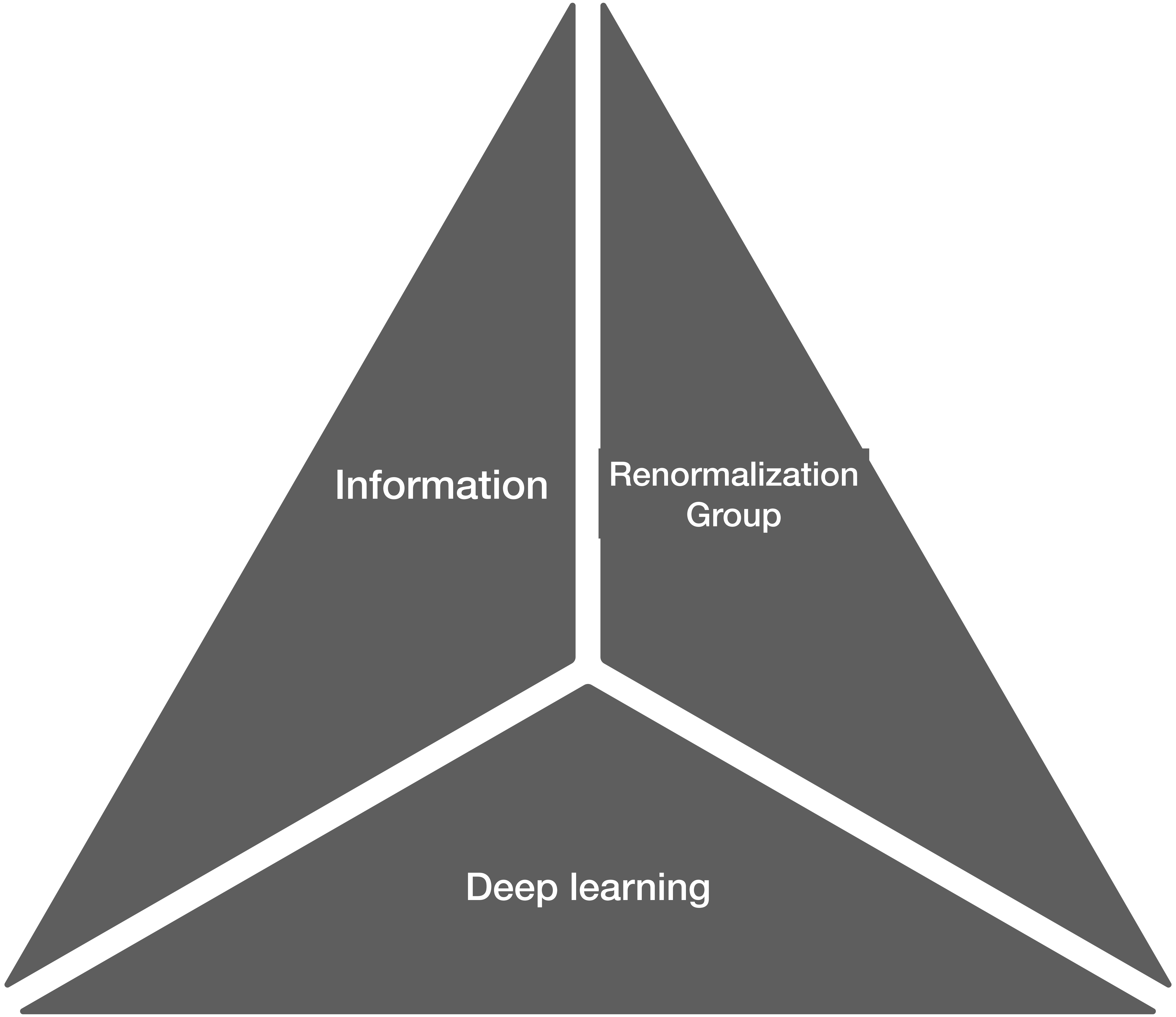
D.E. Gokmen, ZR, S.D. Huber and Maciej Koch Janusz
Physical review letters 127 (24), 240603 (2021)

Snir Gazit

Sounak Biswas

Felix Flicker

Lior Oppenheim


Information
Renormalization Group
Deep learning

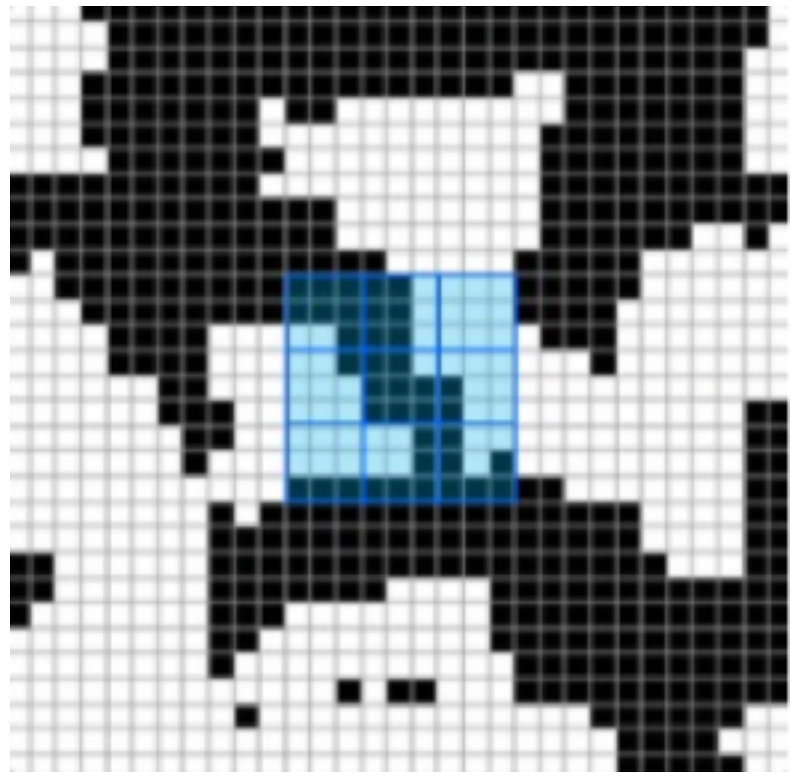
Identification of order-parameters/relevant-operators

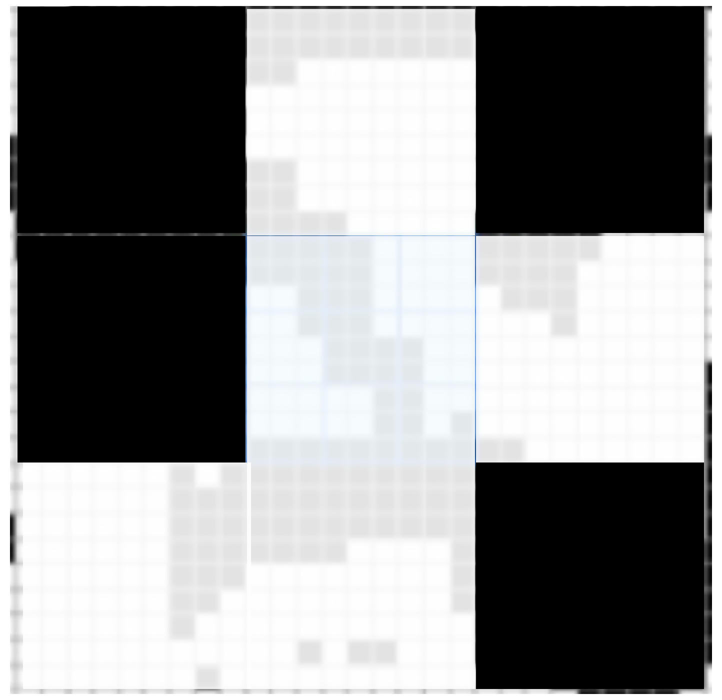
Requires craftsmanship, RG is no silver bullet.

- **When a continuum theory is given:** Do RG: Stable? Good. Unstable? Hmm… Use c-theorems, small-parameters, anomalies, symmetries
- **When only a lattice model is given:** Guess a continuum Lagrangian based on symmetries. Often non-trivial, for instance:
- **Transfer matrix?** Useless for d>2

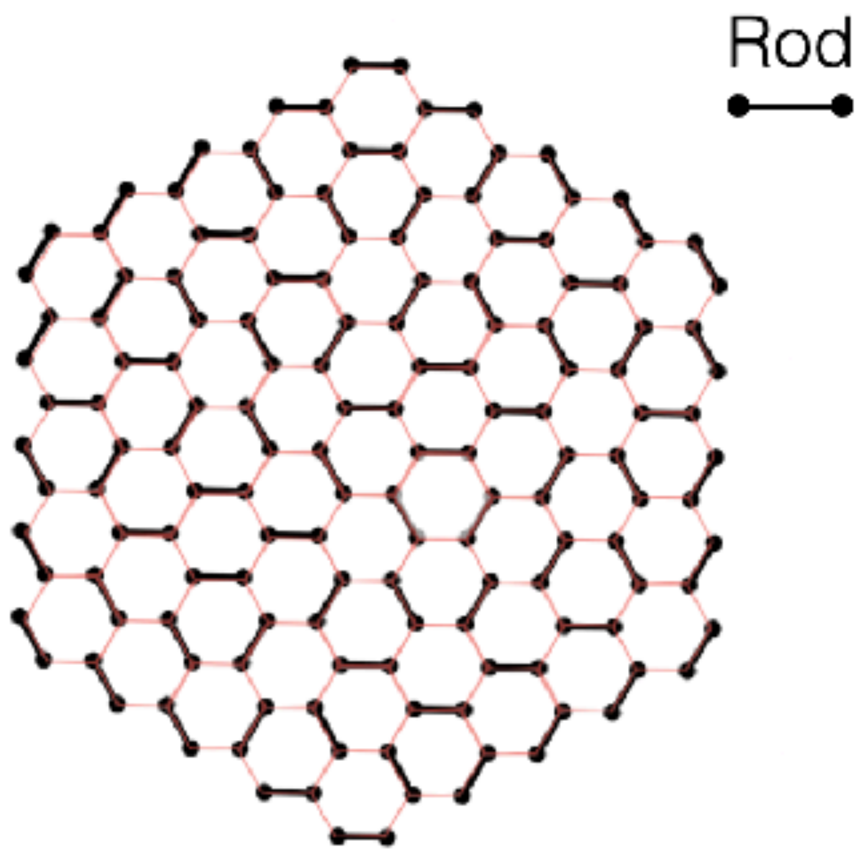
Two illustrative examples

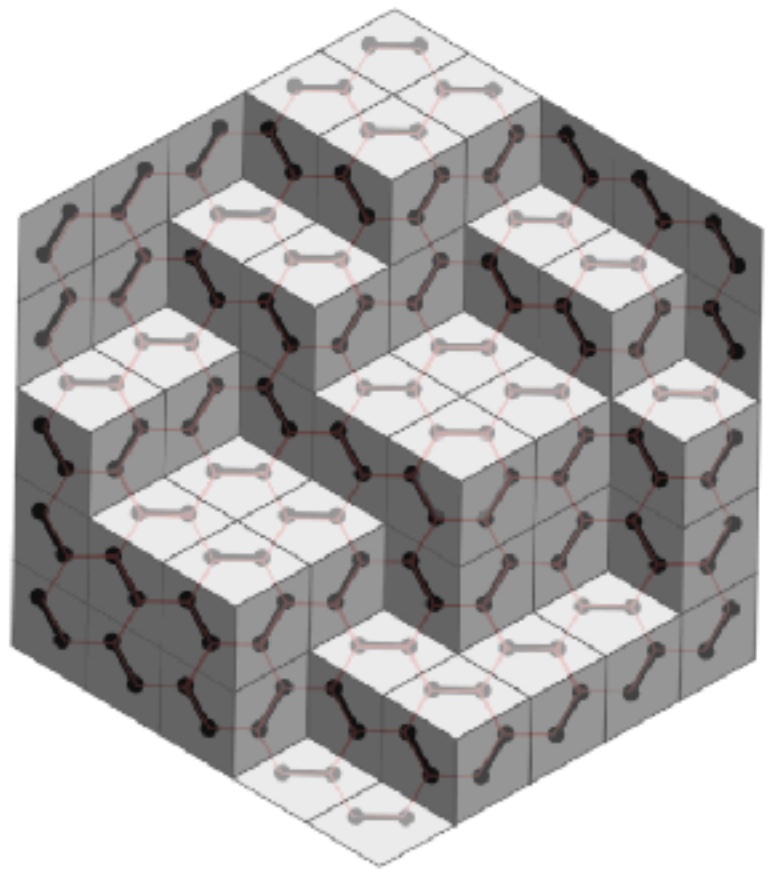
2d Ising Model

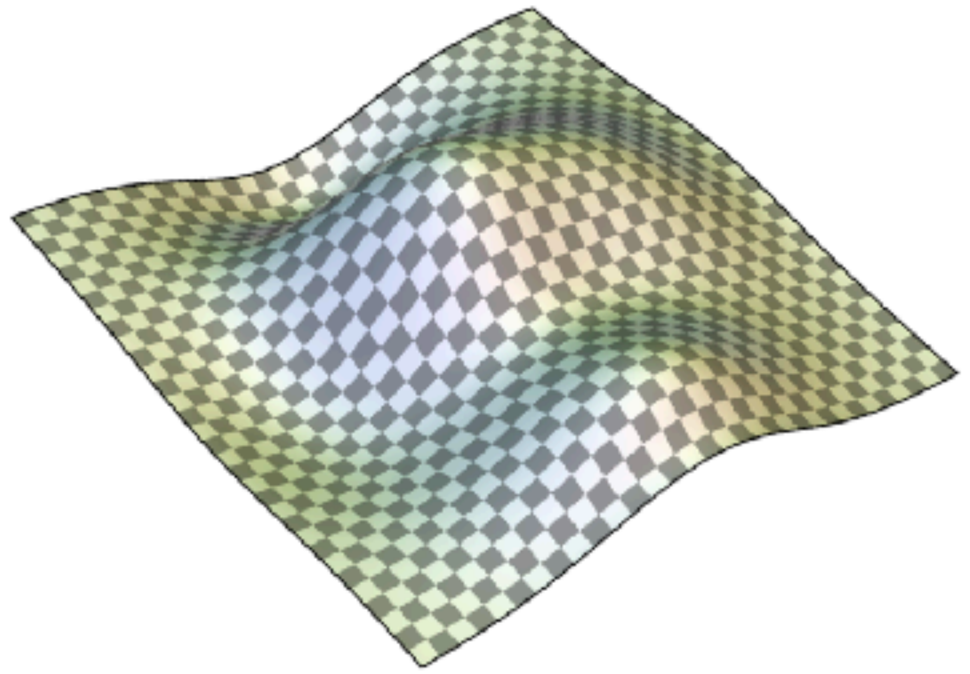




Dimer Model







Order parameters carry the relevant information for explaining low energy experiments

Enter information theory…

Standard Information theory toolbox in physics

These quantify information “without judgement” or notion of relevancy

Entropy,	Von-Neumann entropy,	Entanglement spectrum,	Von-Neumann Mutual Information
Thermodyanmics Boltzmann Eq.	Central charge, Topological degeneracies	Topological phases of matter, DMRG numerics	C-theorems in d=2,3,4 [Huerta,Cassini]

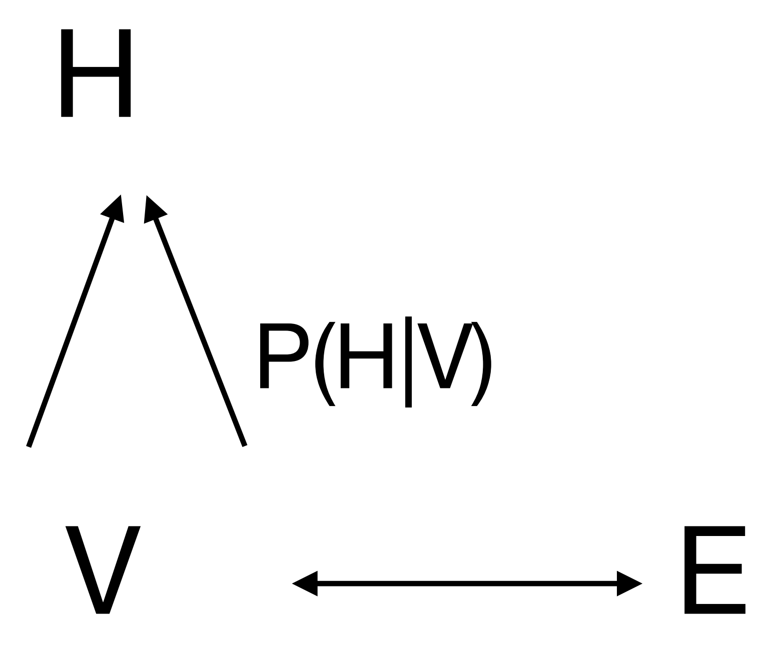
More recently (2000) a way of identifying relevant-information has been proposed - the Information Bottleneck

The Information Bottleneck (IB)

How to define the best lossy compression of “relevant” information

Consider compressing one random variable (V) into another (H) when what you care about is information on a third one (E).

A solution is some conditional probability P(H|V)



The IB approach seeks P(H|V) that minimizes:

$$\min_{p(H|V)} \mathcal{L}[p(H|V)] = I(H,V) - \beta I(H,E)$$

$\boldsymbol{\beta}$ is not to be confused with physical temperature!

Strong compression → Strong knowledge on E

β

N. Tishby, F.C. Pereira, W. Bialek (2000)

The Information Bottleneck (IB) - Numerical aspects

Unfortunately, getting the optimal $p(\mathrm{H}|\mathrm{V})$ in a generic setting is an NP-hard problem.

On the bright side, the same can be said about the 3d Ising model.

Recent advancements in deep learning make this numerically tractable.

$$\min_{p(H|V)} \mathcal{L}[p(H|V)] = I(H,V) - \beta I(H,E)$$

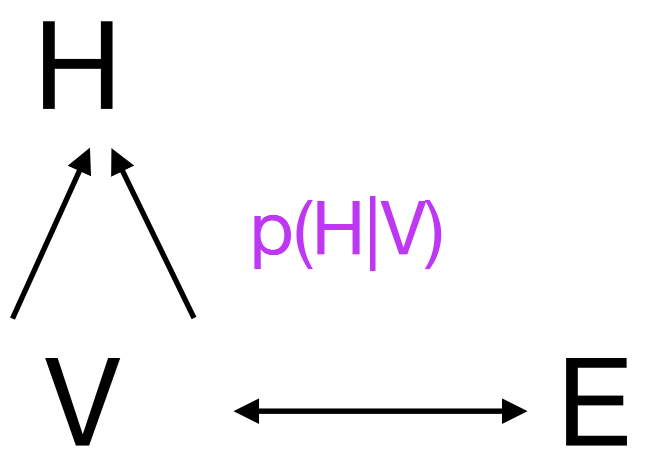
Strong compression → Strong knowledge on E

β

IB transitions/bifurcations

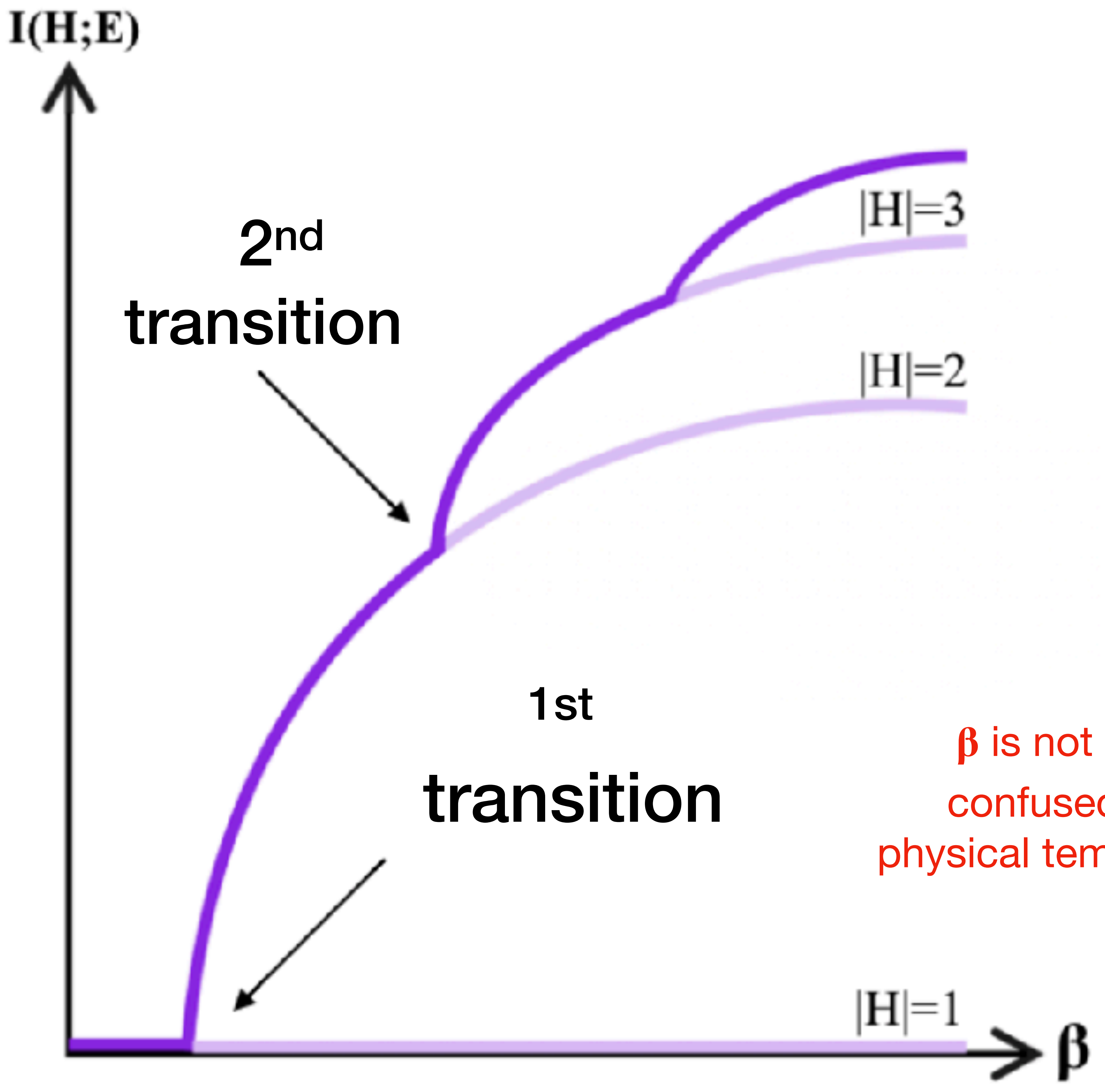
The moment a new feature is tracked

$$\min_{p(H|V)} \mathcal{L}[p(H|V)] = I(H,V) - \beta I(H,E)$$



Strong compression — Strong knowledge on E

β →



β is not to be confused with physical temperature!

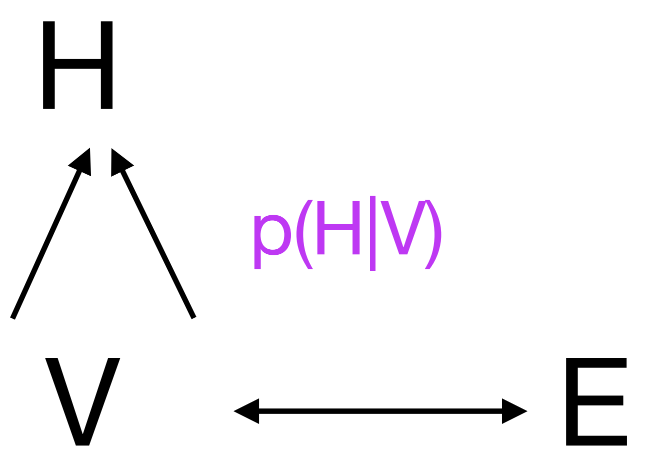
Can be viewed as symmetry breaking transitions of the permutation symmetry of L_{IB} w.r.t. to the elements in H - Many nice analogies with Landau theory of phase transitions

[Gedeon, Parker, Dimitrov (2012)]

IB transitions/bifurcations

The moment a new feature is tracked

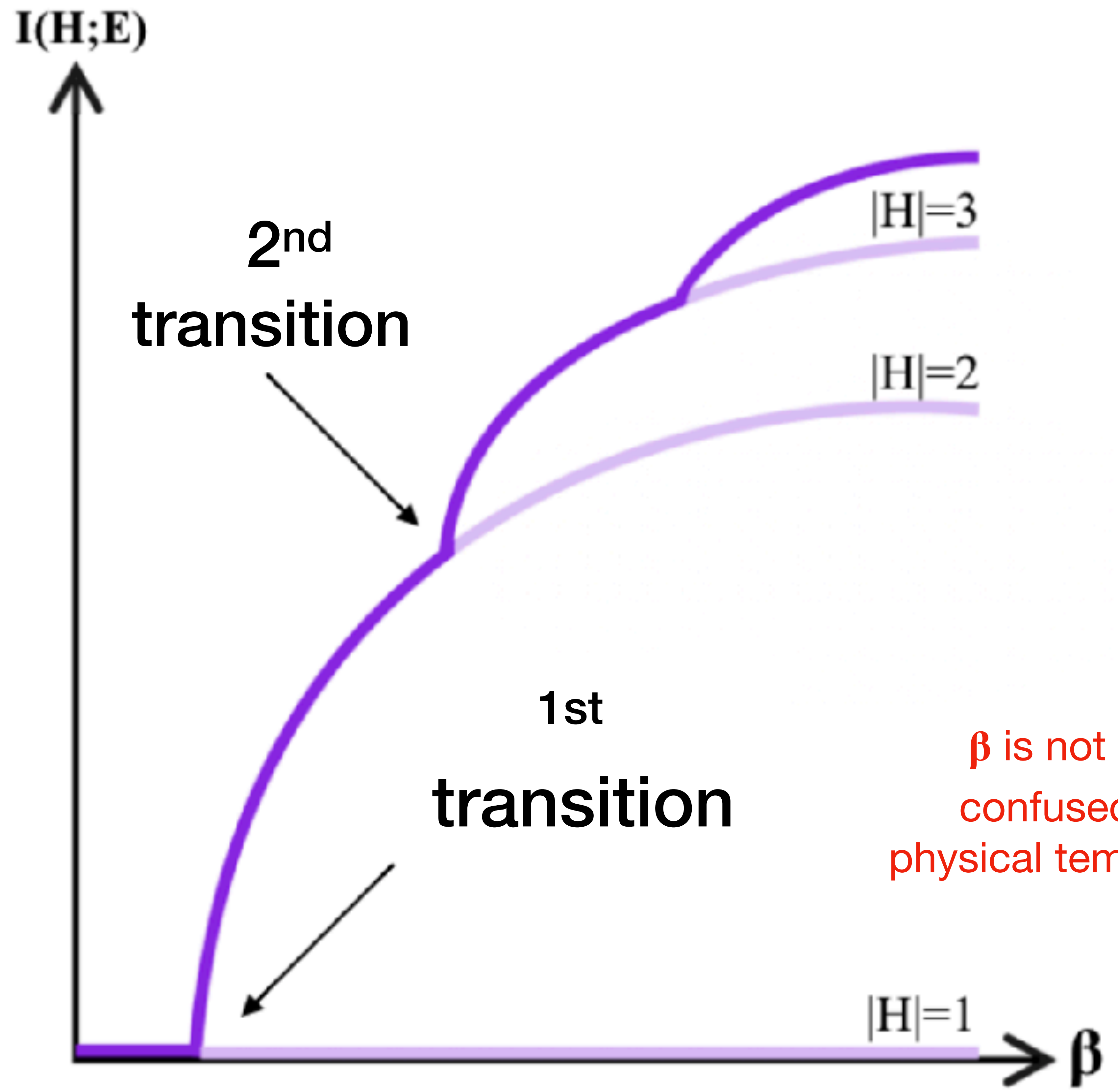
$$\min_{p(H|V)} \mathcal{L}[p(H|V)] = I(H,V) - \beta I(H,E)$$



Strong compression — Strong knowledge on E

$|H|>>1$ ⟶ β

$\beta >>1$ ⟶ $|H|^{-1}$



β is not to be confused with physical temperature!

Can be viewed as symmetry breaking transitions of the permutation symmetry of L_{IB} w.r.t. to the elements in H - Many nice analogies with Landau theory of phase transitions

[Gedeon, Parker, Dimitrov (2012)]

Conjectured IB-relevant-operators relation (RSMI)

$$\min_{p(H|V)} \mathcal{L}[p(H|V)] = I(H,V) - \beta I(H,E)$$

H

p(H|V)

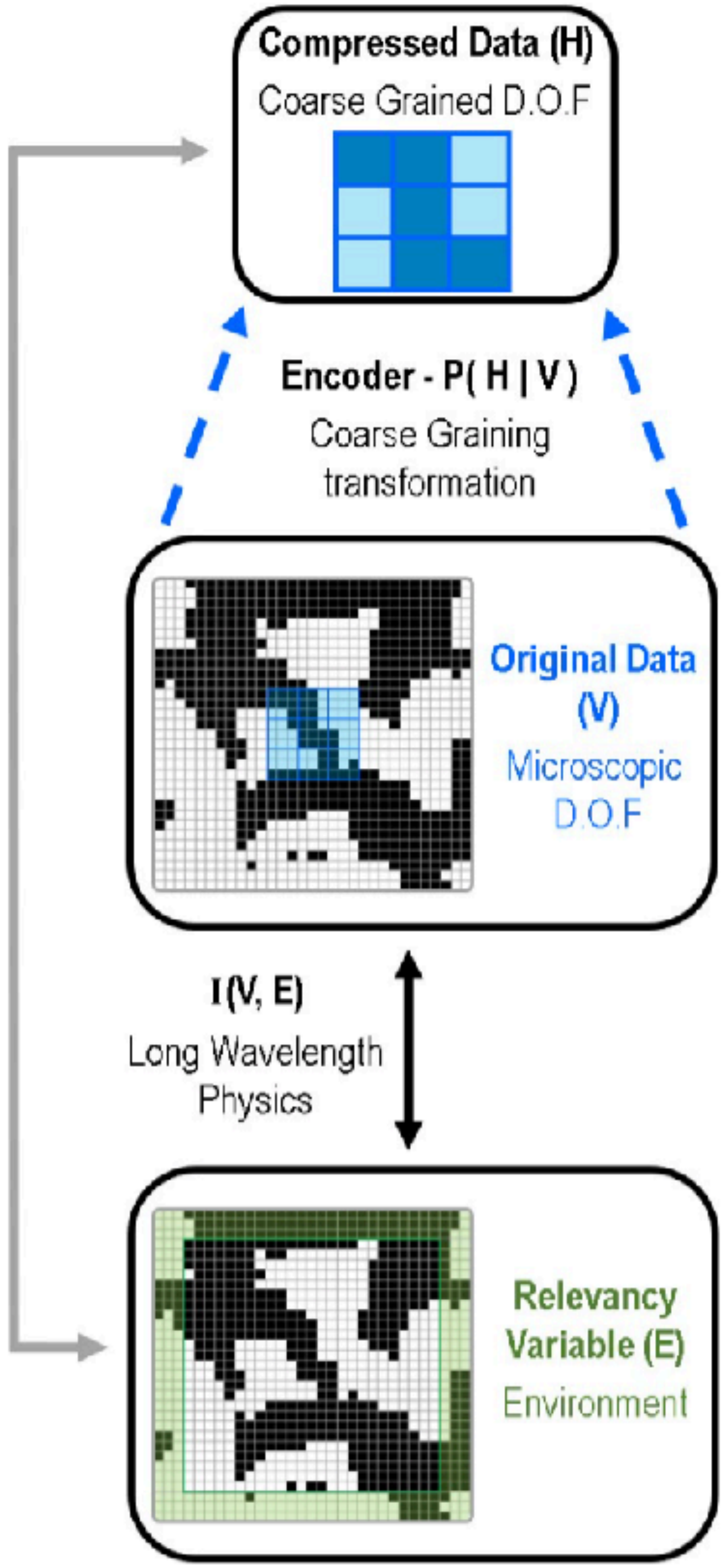
Should encode relevant operators

V ⟷ E

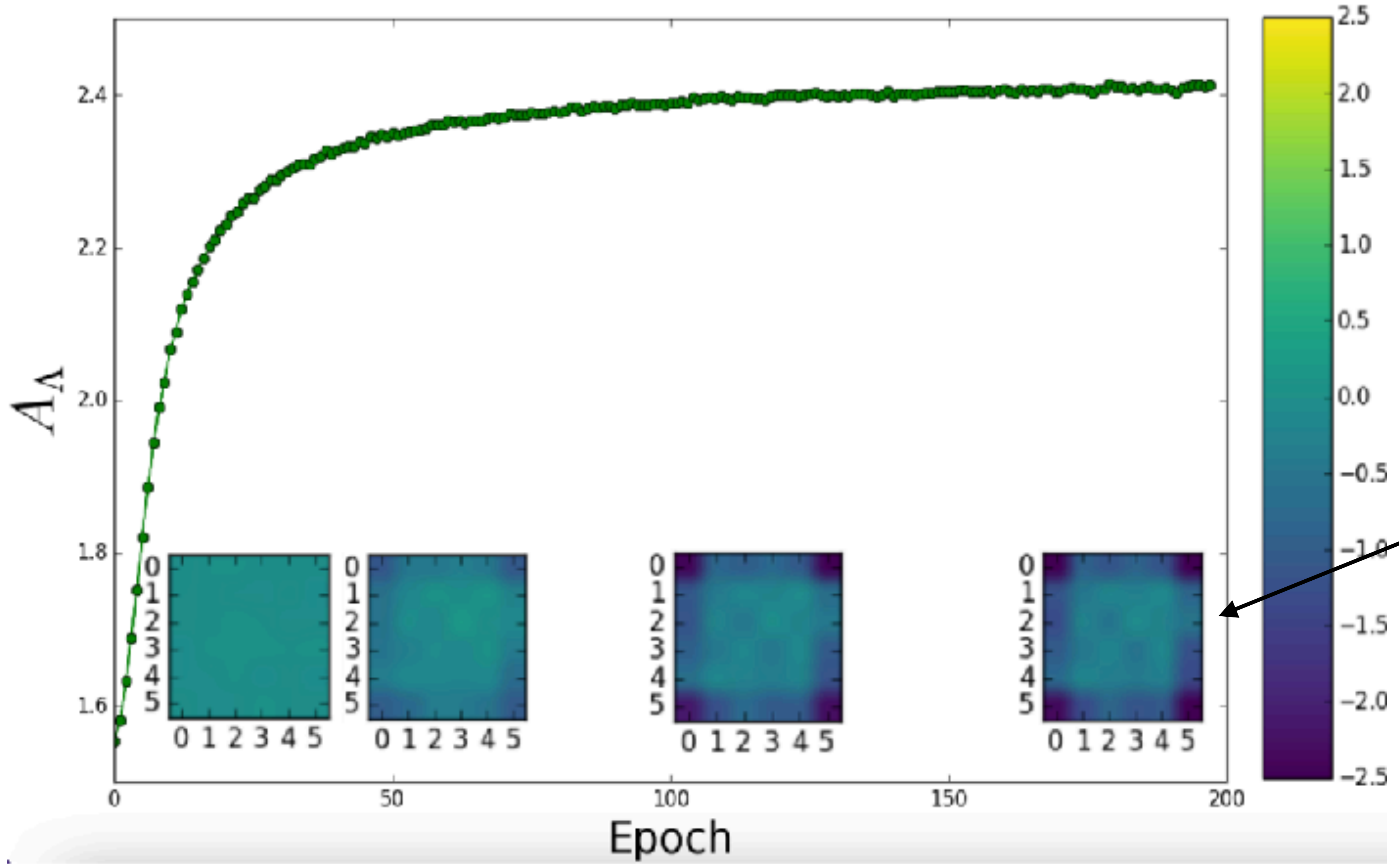
Strong compression — Strong knowledge on E

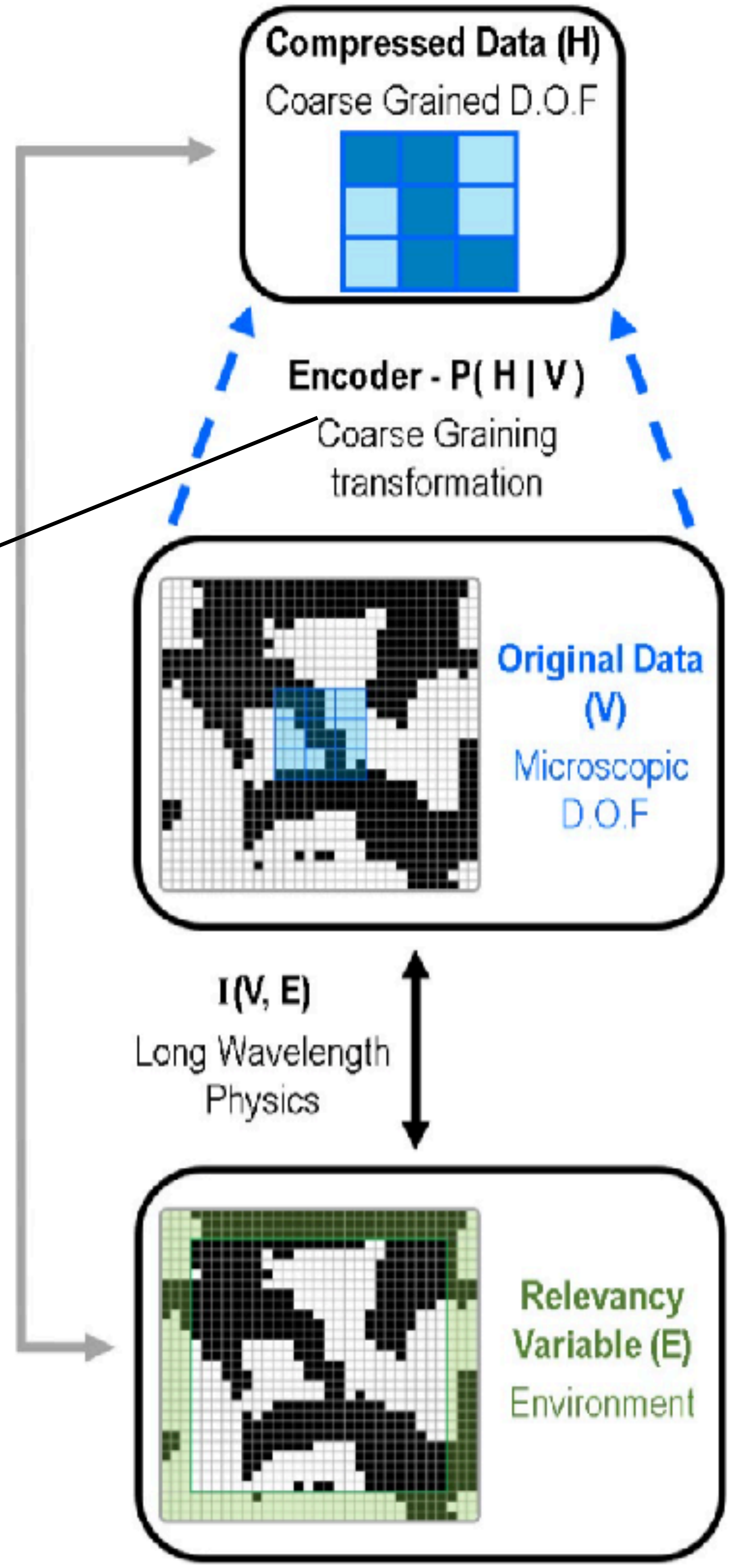
|H|>>1 ⟶ β

β >>1 ⟶ $|H|^{-1}$



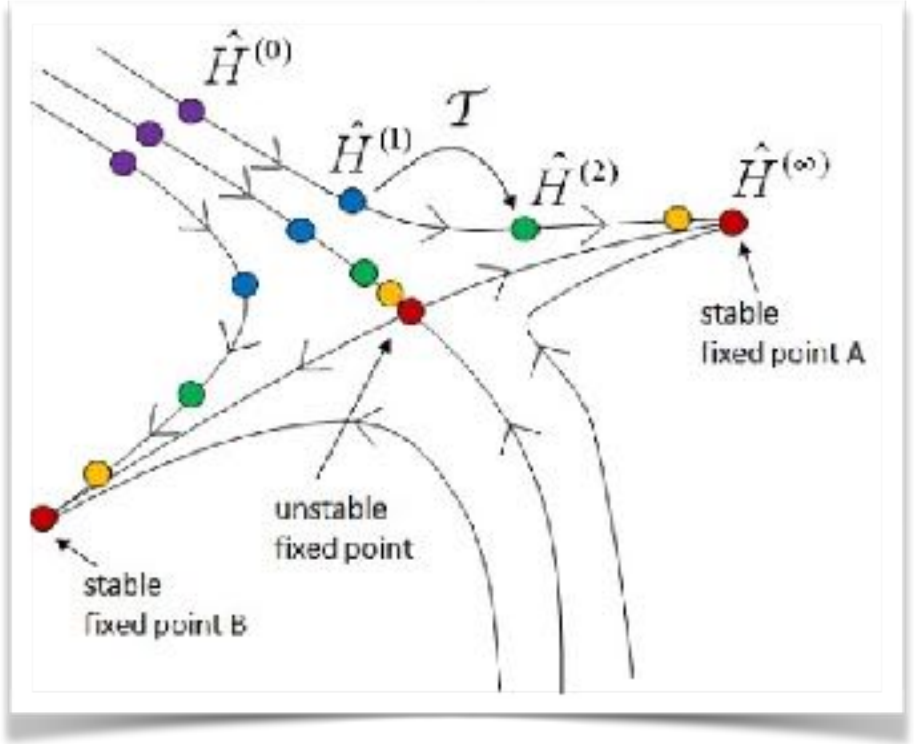
Maciej Koch Janusz, and ZR, *Nature Physics 14, 578-582 (2018)*





How is this related to field theory?

Equivalent notions of relevance in physics



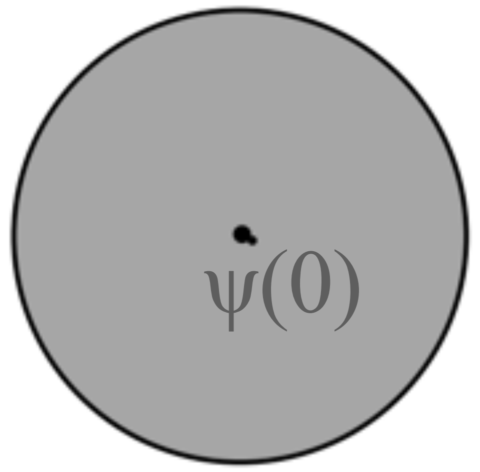
Renormalisation Group

Adding $\psi(x)$ leads to an increase of its coupling constant under the RG flow

CFT

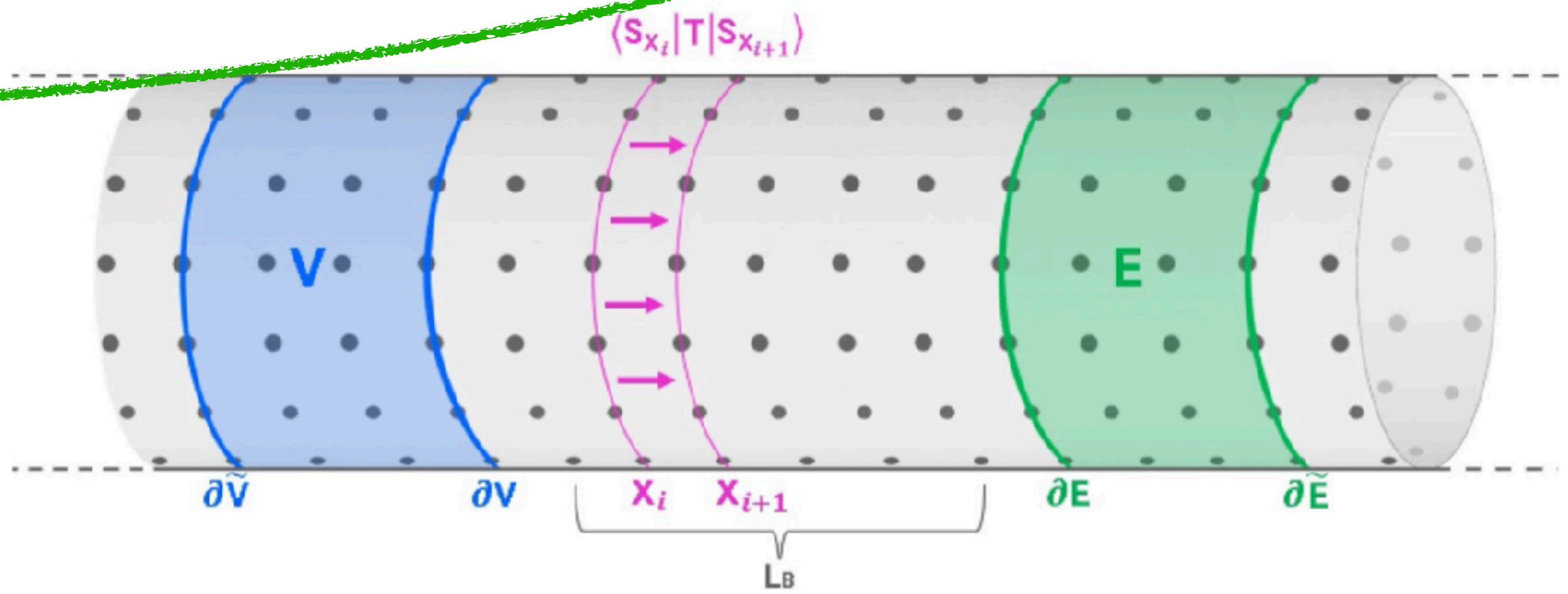
$\psi(x)$ and $\psi(y)$ correlations decay as a power law with a small exponent (2Δ)

In radial quantisation $\psi(0)$ generates a eigenstate of the scaling operator with large eigenvalue (Δ)

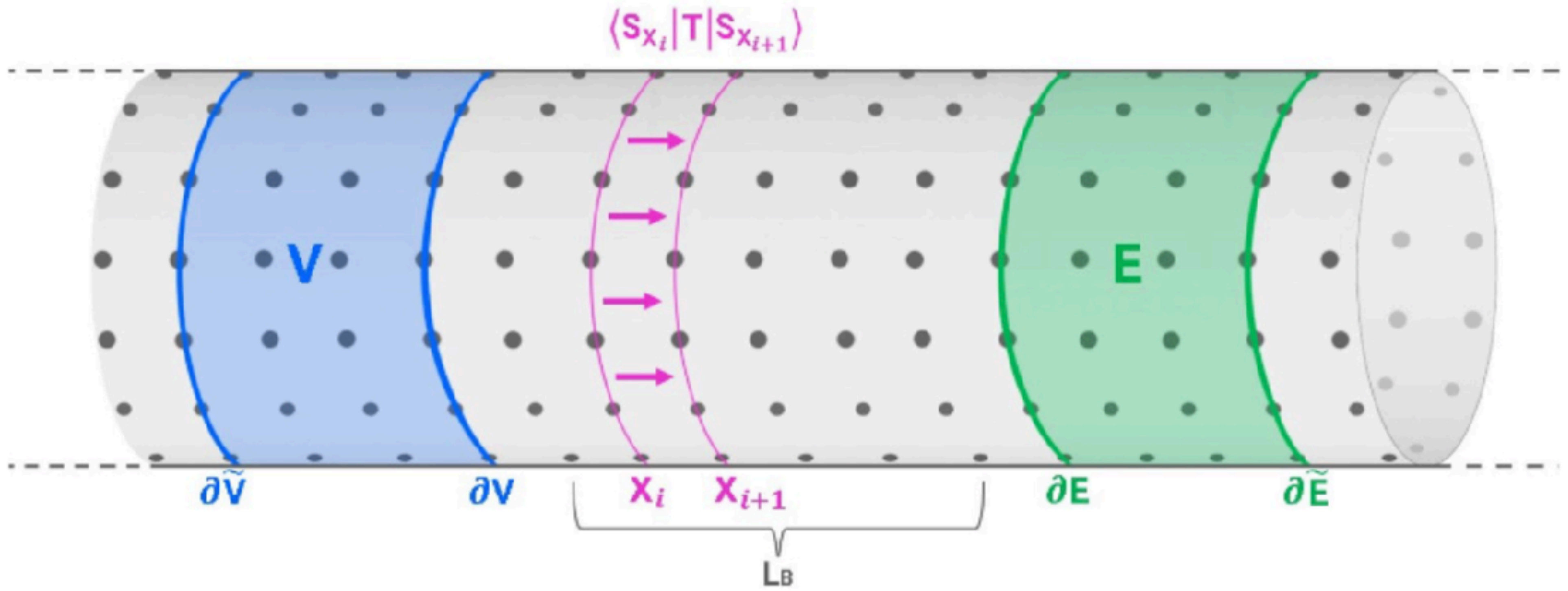


Transfer-matrix (T)

On a cylinder, given $T|0\rangle = |0\rangle$ is the maximal eig. of T, $T|\psi\rangle = T\psi|0\rangle = e^{-4\pi\Delta/L}|\psi\rangle$



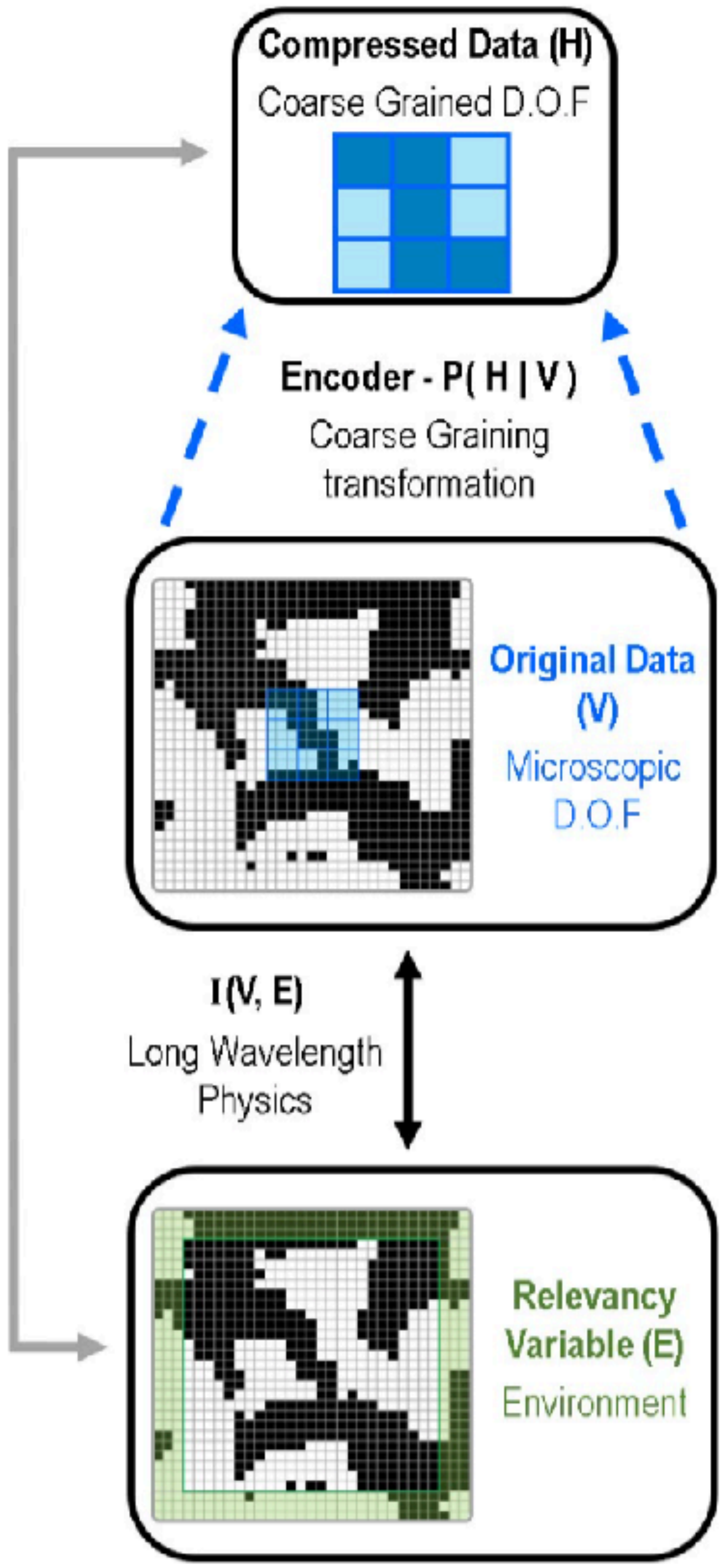
Transfer matrix reminder

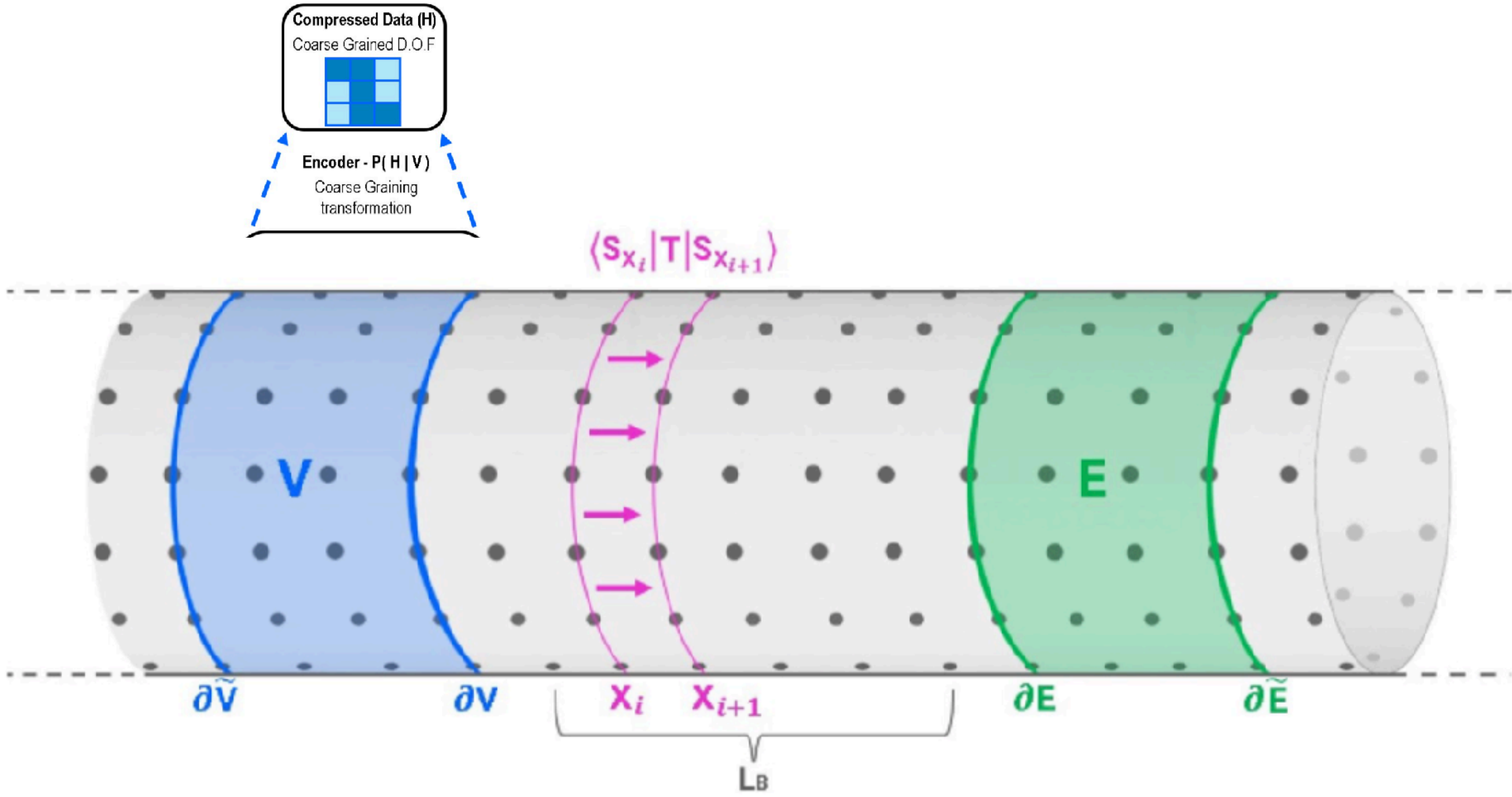


- Reflects the Boltzmann-factors, on and in between two consecutive ("time"-)slices
- Yields the partition function via $Z = \mathrm{Tr}[T^L]$
- Element-wise positive (also in many quantum problems)
- Columns sum to 1 in Markov Chain case
- Large eigenvalues/eigenvectors closely related with CFT's notion of relevance:

$T|0\rangle = |0\rangle$ is the maximal eig. of T,
$T|1\rangle = e^{-4\pi\Delta/L}|1\rangle$ is the sub-leading eig. of T,
$|1\rangle = \Psi\,|0\rangle$ Ψ is the leading primary operator

Linking IB and transfer matrix
Step 1 - move our IB machinery to the cylinder

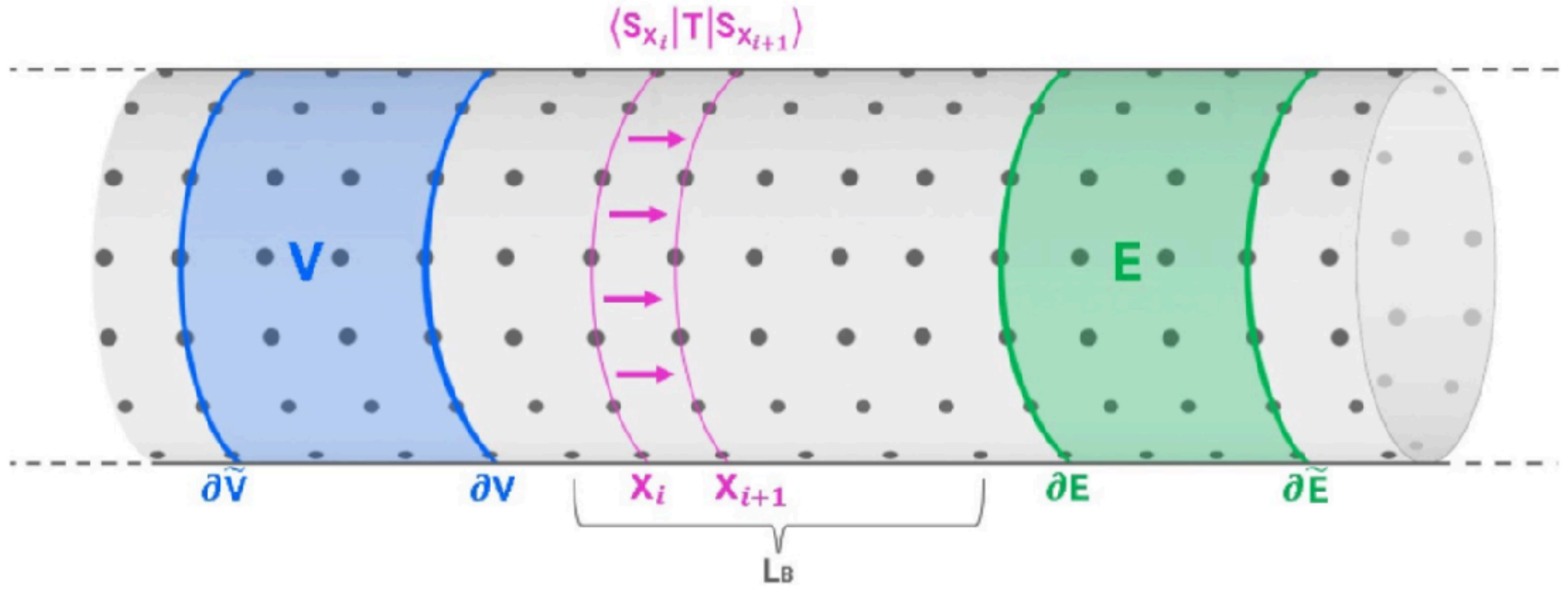




*This geometry is also easier to relate to Markov Chains, with —> time —>

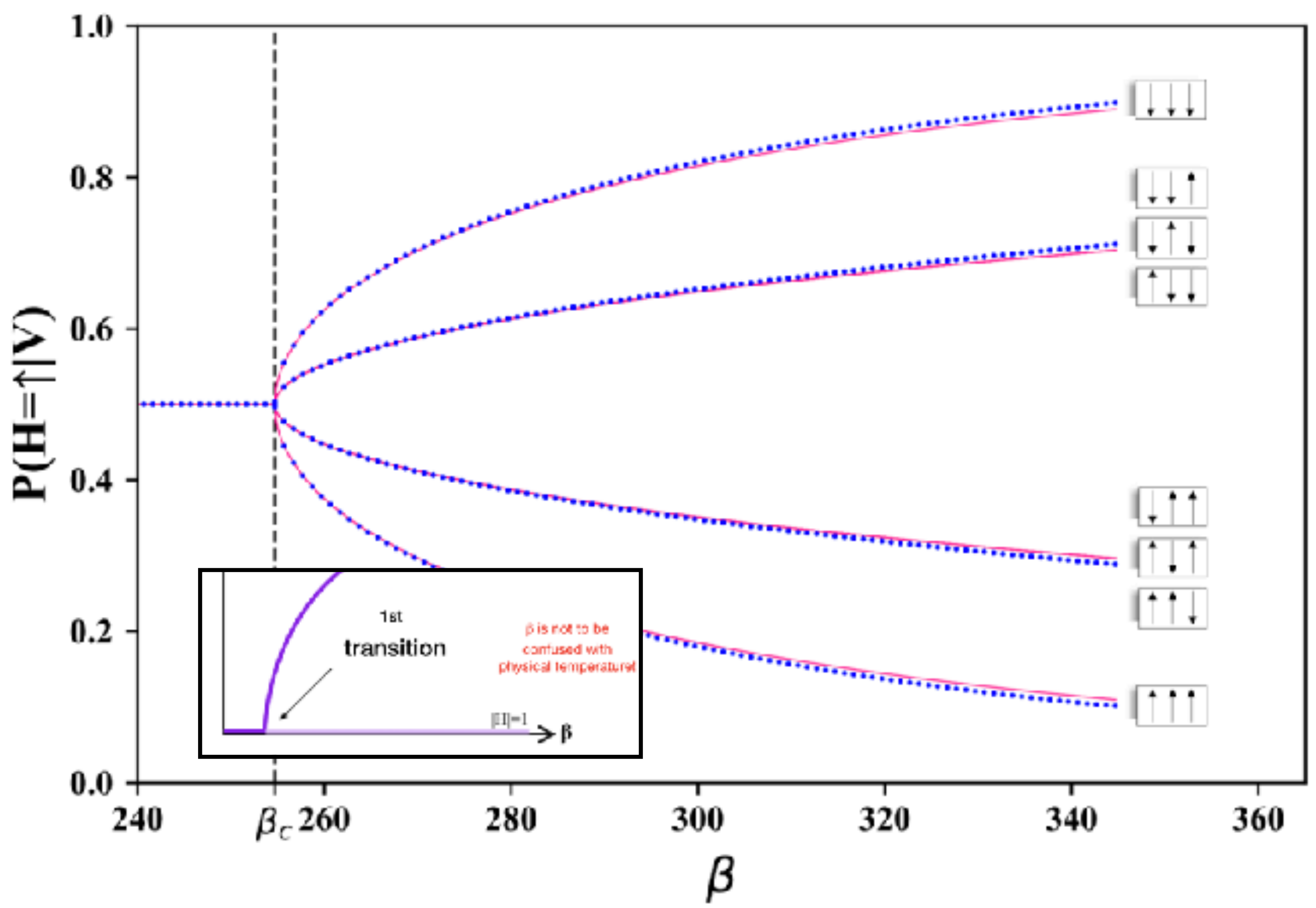
Linking IB and Transfer-Matrix
Step 2: Solving IB analytically in terms of transfer-matrix data

$$r_v = \frac{\langle 1|\partial V\rangle}{\langle 0|\partial V\rangle} = \frac{\langle 0|\psi|\partial V\rangle}{\langle 0|\partial V\rangle}$$



$$p(h=\pm 1|v) = \frac{e^{hm(t)r_v}}{2\cosh(m(t)r_v)}$$

"Critical" Ising on an L_c=3 cylinder



$$m(t) = \sqrt{\frac{3(\beta-\beta_{c1})}{\langle r_x^4\rangle \beta_{c1}}} \quad \beta_{c1} = e^{8\pi\Delta_1 \frac{L_B}{L}}$$

Amit Gordon, Aditya Banerjee, Maciej Koch Janusz, and ZR PRL (2021)

A closer look at r_v

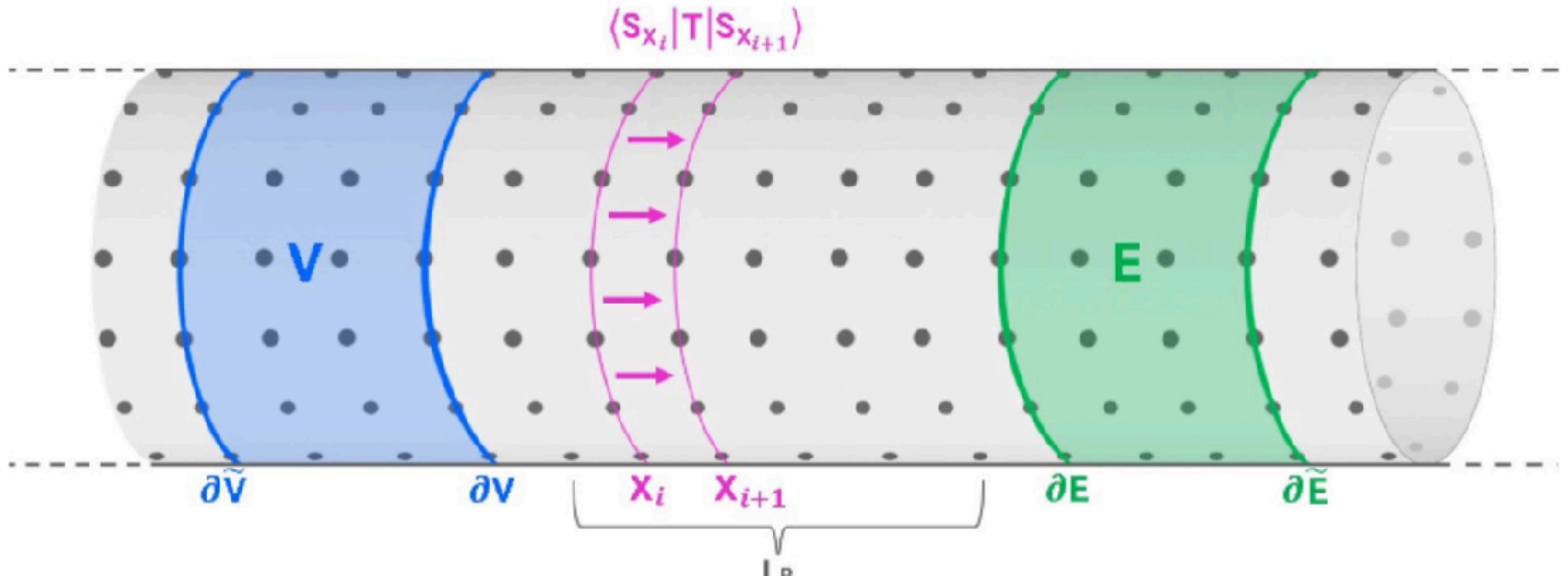
$$r_v = \frac{\langle 1|\partial V\rangle}{\langle 0|\partial V\rangle} = \frac{\langle 0|\psi|\partial V\rangle}{\langle 0|\partial V\rangle}$$

2d Ising model → Average magnetisation of ∂V

Primary Op is diagonal → Diagonal entry corresponding to ∂V

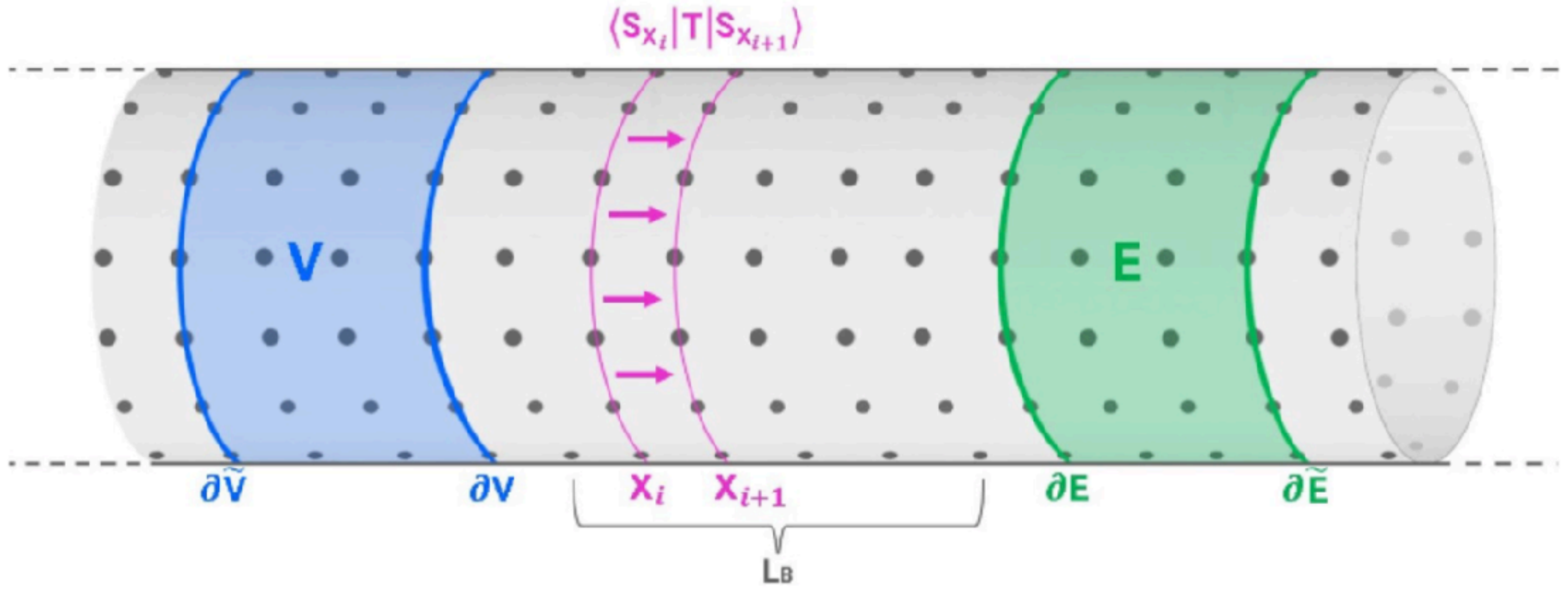
Markov Chain → $\langle 1|\partial V\rangle$

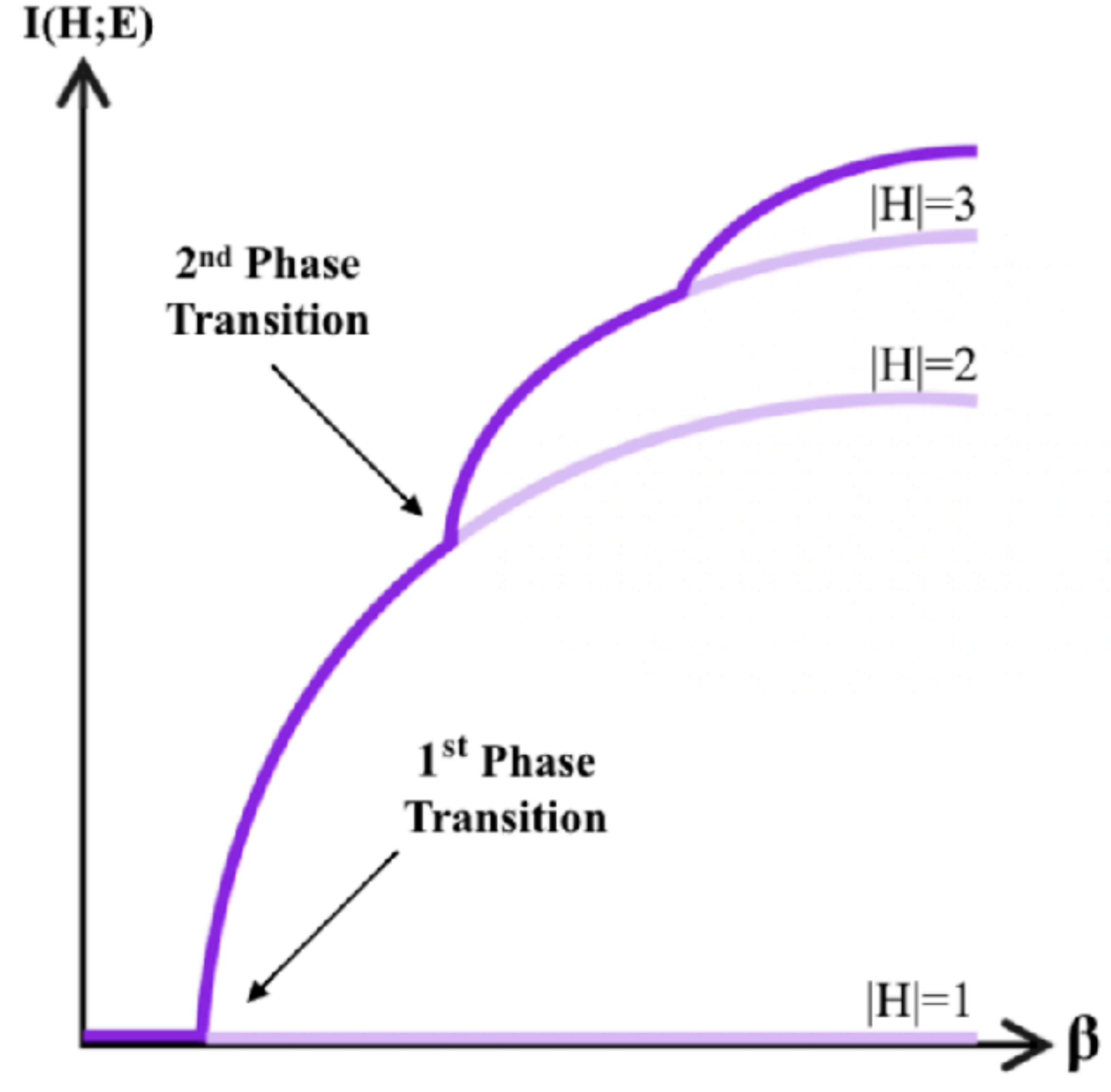
Average = 0
Var = 1



IB - RG dictionary

If IB is a microscope, how does it optics work?





- The first feature IB tracks is the “normalized” coefficients of the sub-leading eigenvector of the transfer matrix.
- The first critical β gives the scaling dimension of that operator $\beta_{c1} = e^{8\pi\Delta_1 \frac{L_B}{L}}$
- In any local theory, where a transfer-matrix can be defined, IB disregards the bulk of V.
- **A means of accessing transfer-matrix eigenvectors (in any dimension) from Monte-Carlo snapshots!**

Amit Gordon, Aditya Banerjee, Maciej Koch Janusz, and ZR PRL (2021)

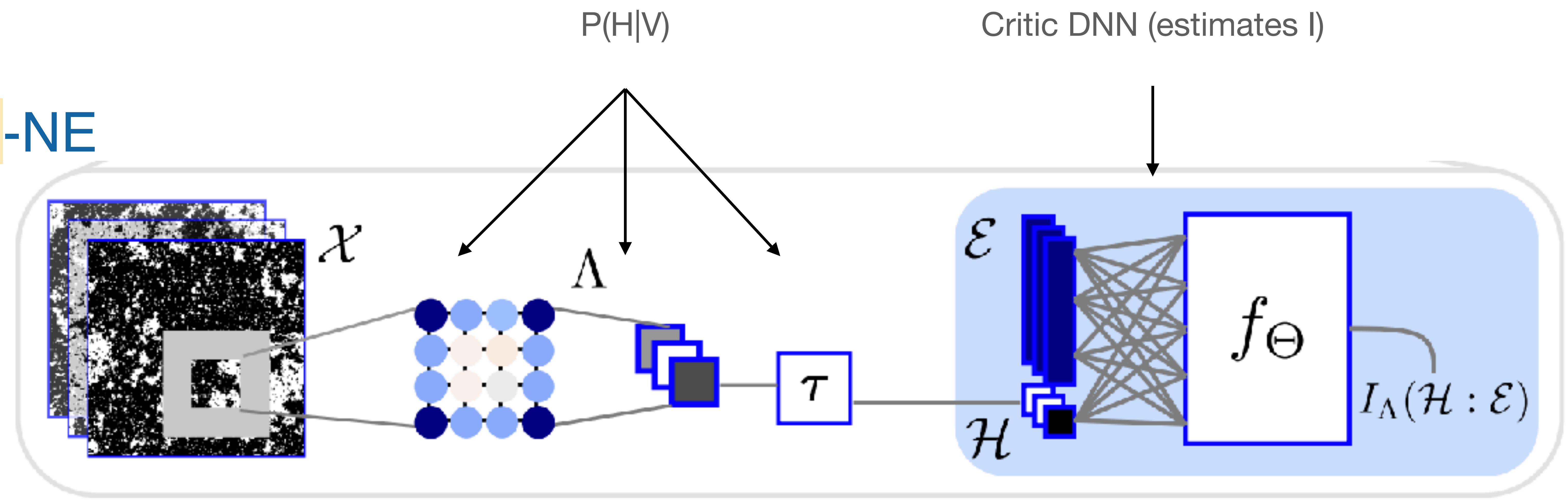
Mid lecture summary

- A concrete dictionary is emerging between information theory notion of relevance (IB) and physics notion of relevance (large-transfer-matrix eigenvectors/relevant-operators)
- Several deep learning based tools allow controlled numerically solutions of IB

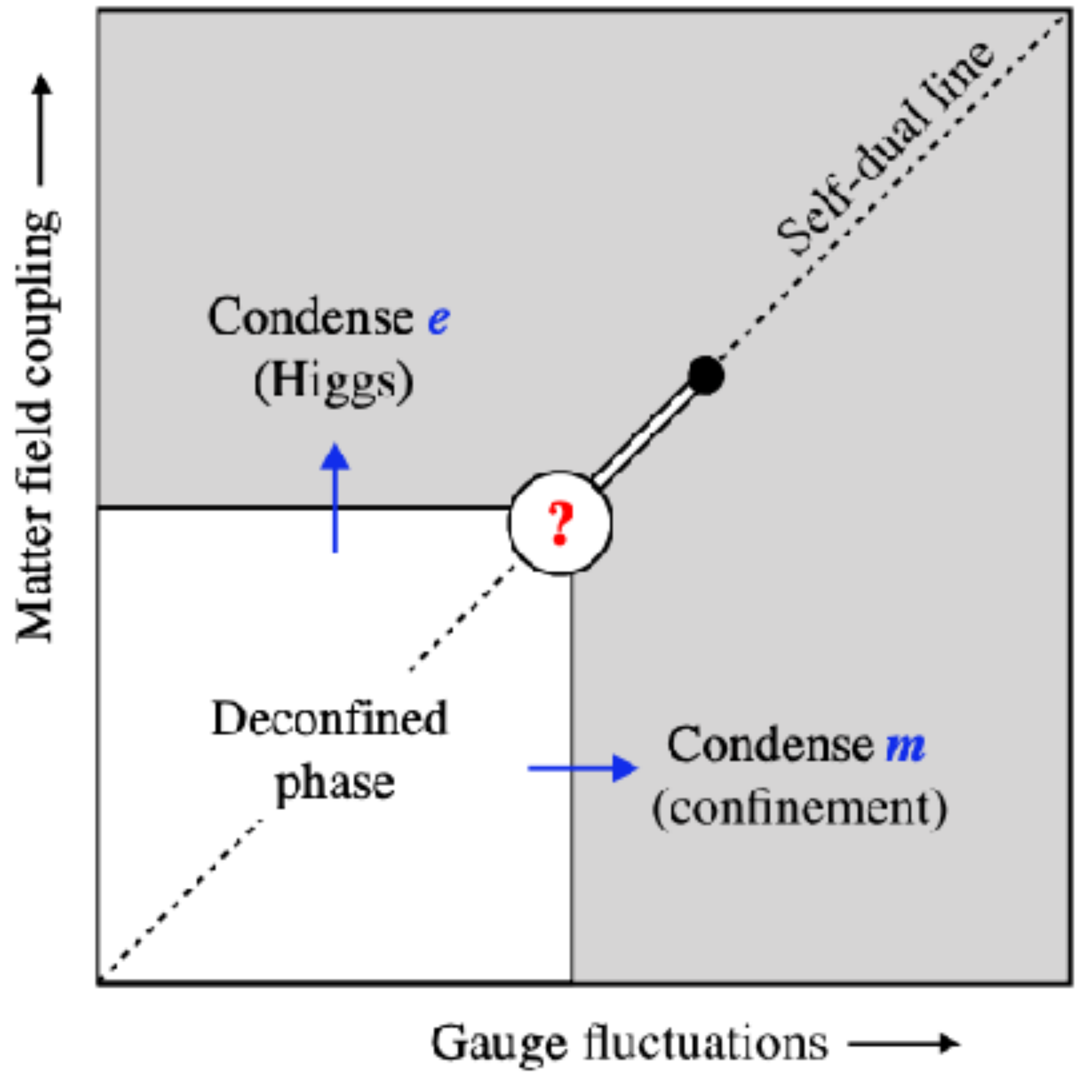
Statistical Physics through the Lens of Real-Space Mutual Information

Doruk Efe Gökmen, Zohar Ringel, Sebastian D. Huber, and Maciej Koch-Janusz
Phys. Rev. Lett. **127**, 240603 – Published 6 December 2021

https://github.com/RSMI-NE/RSMI-NE



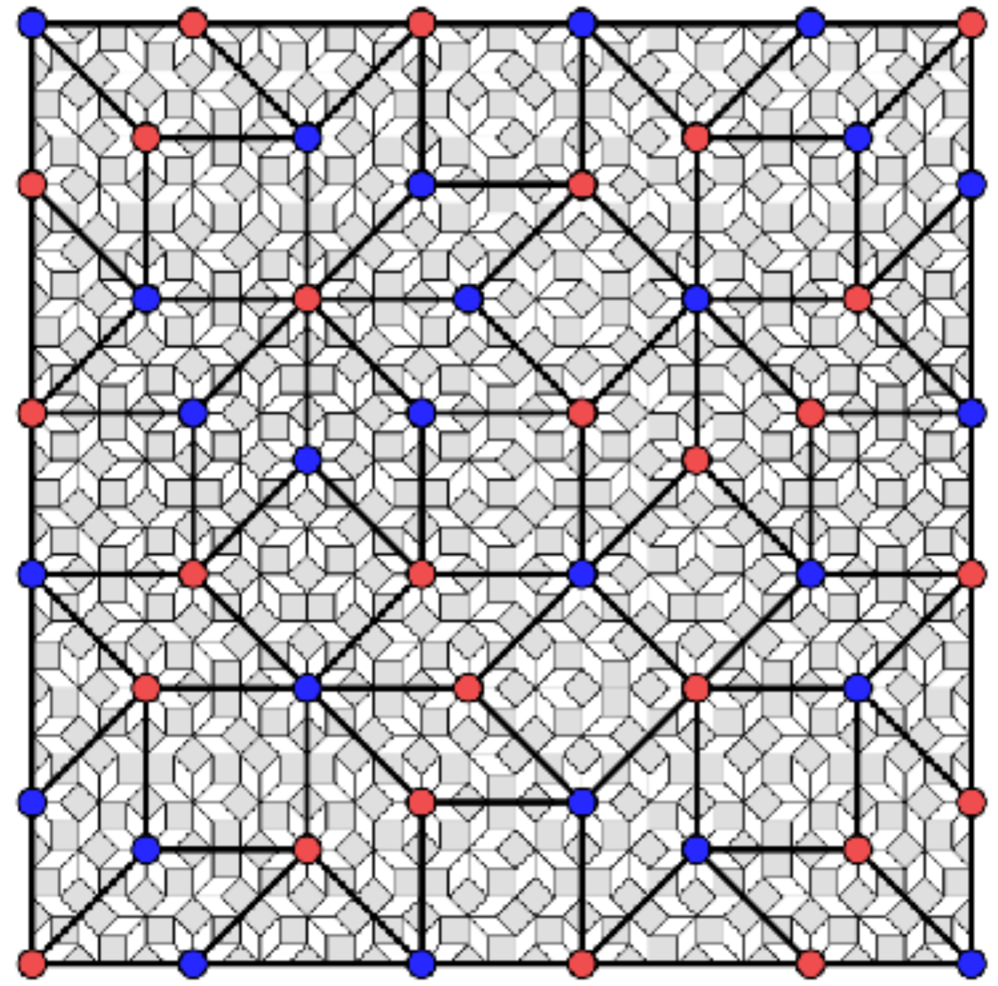
Venturing into the unsolved



PHYSICAL REVIEW X 11, 041008 (2021)

Self-Dual Criticality in Three-Dimensional $\mathbb{Z}_2$ Gauge Theory with Matter

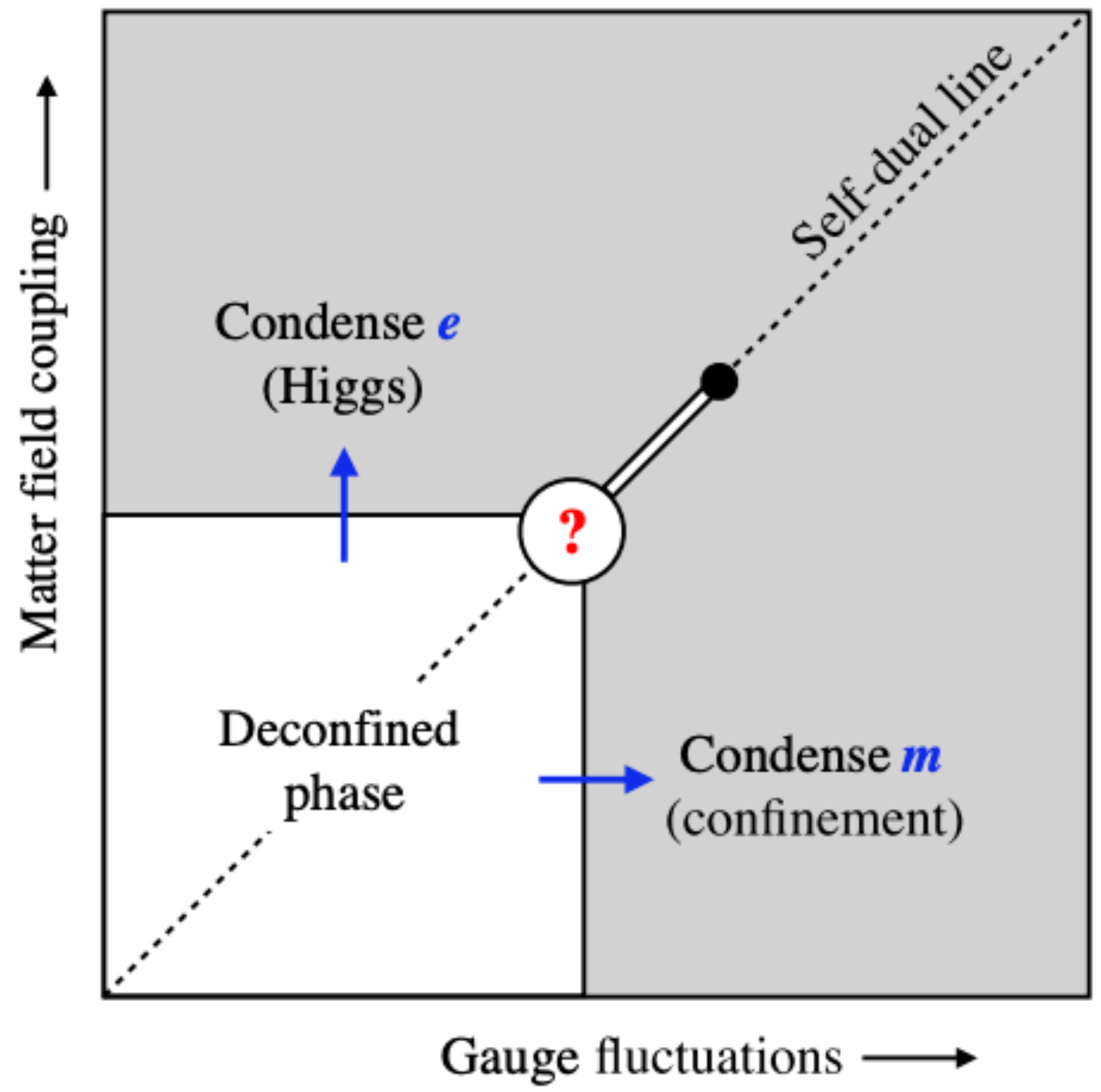
Andrés M. Somoza,[1] Pablo Serna,[1,2] and Adam Nahum[2,3]



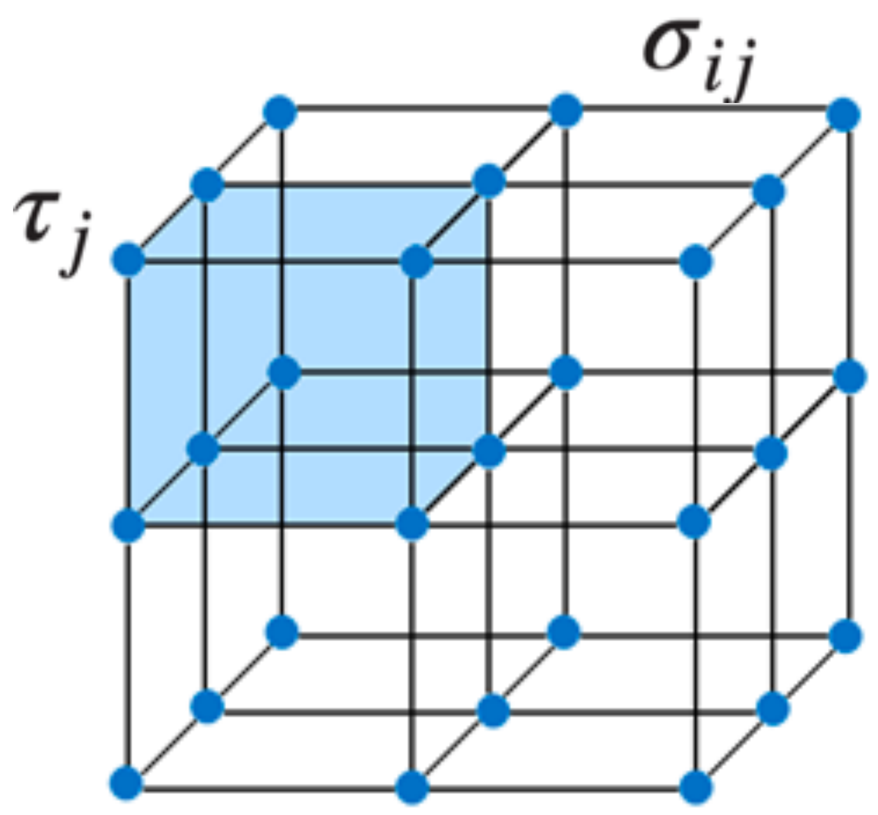
Statistical mechanics of dimers on quasiperiodic tilings

Jerome Lloyd,[1,2,3,*] Sounak Biswas,[1,*] Steven H. Simon,[1] S. A. Parameswaran,[1] and Felix Flicker[1,4]

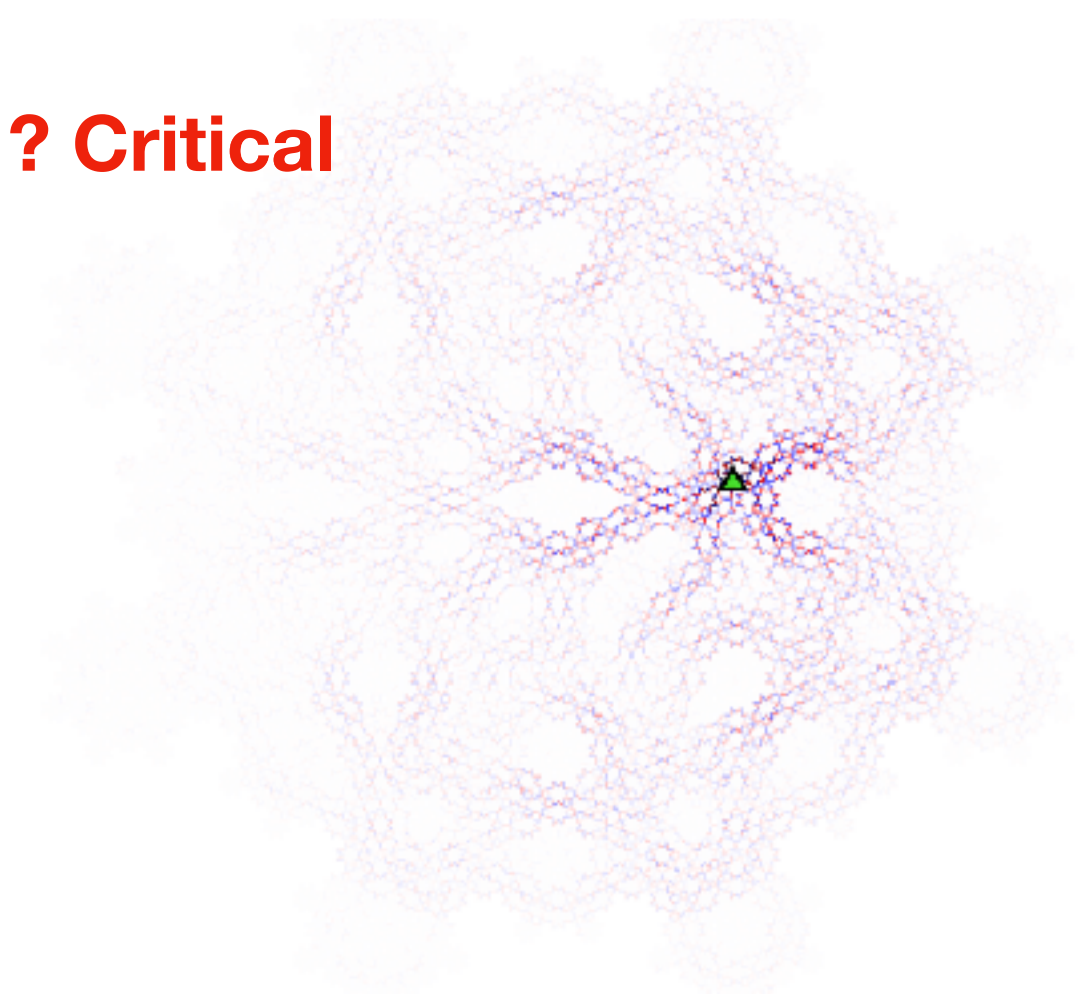
PRB 2022

- 

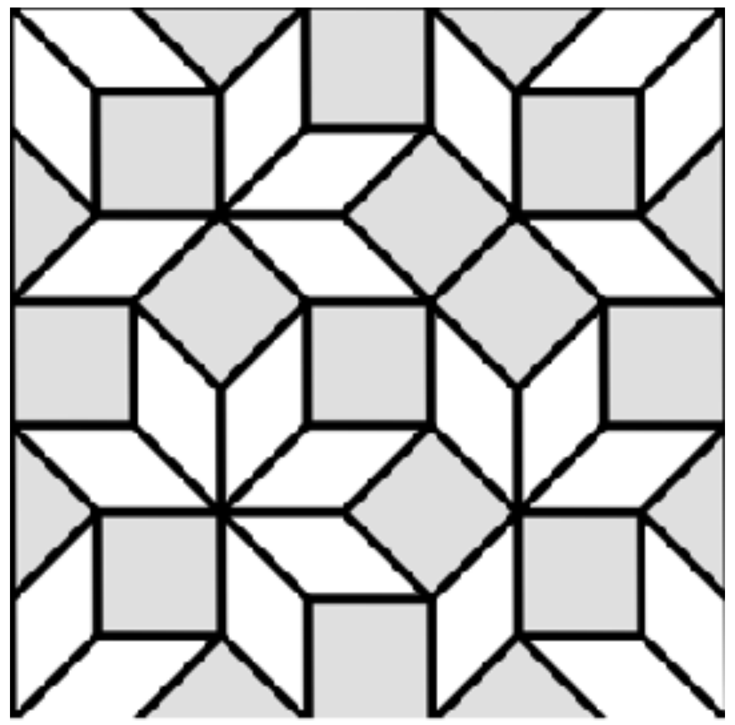
$$Z \propto \sum_{\{\sigma\},\{\tau\}} \exp\left(K\sum_{\square}\left(\prod_{\langle ij\rangle\in\square}\sigma_{ij}\right) + J\sum_{\langle ij\rangle}\tau_i\sigma_{ij}\tau_j\right)$$



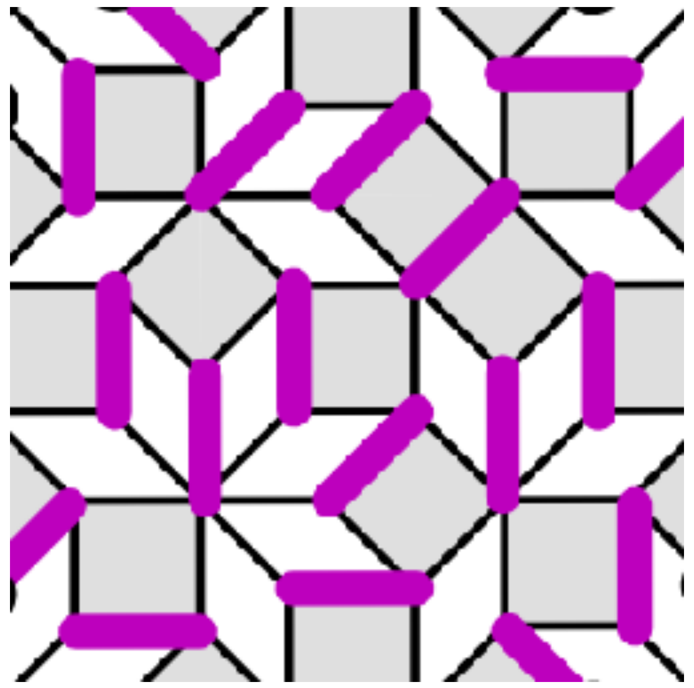
- ? Critical



Amman Beenaker Tiling

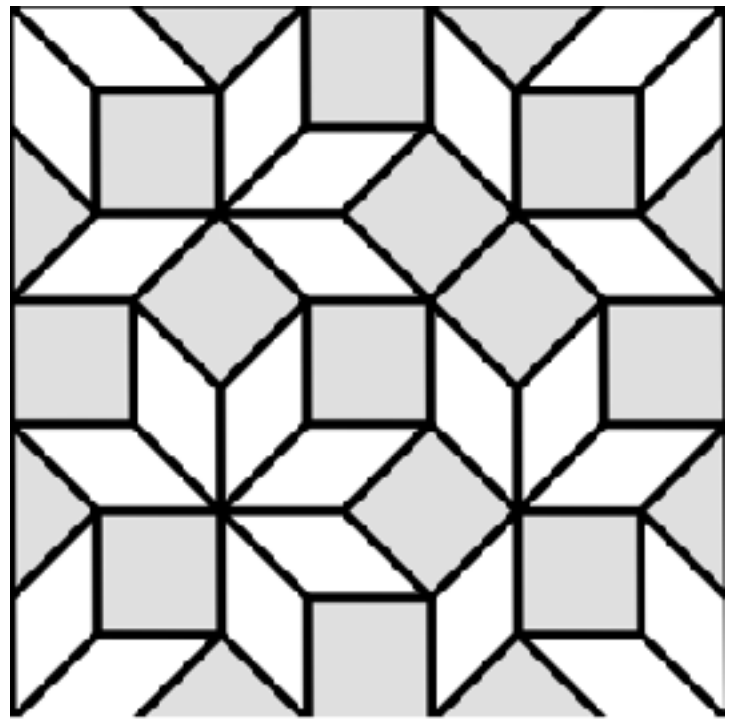


With dimers

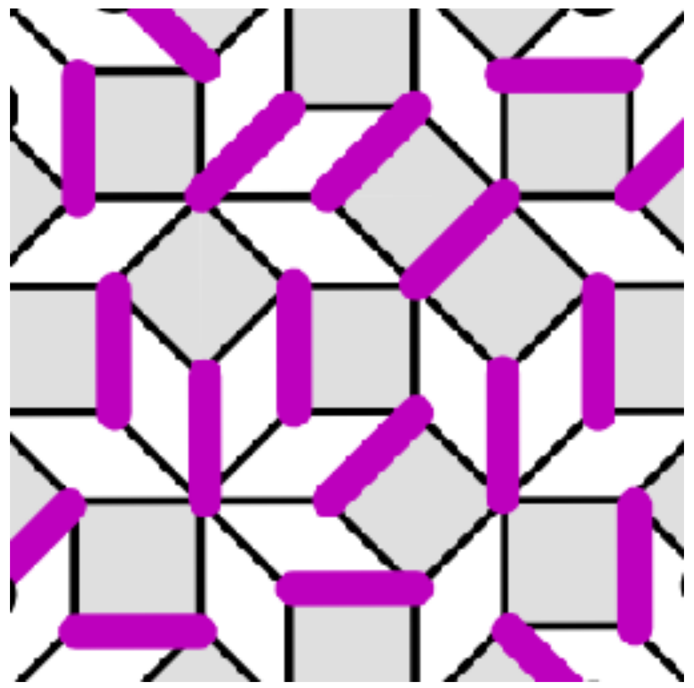


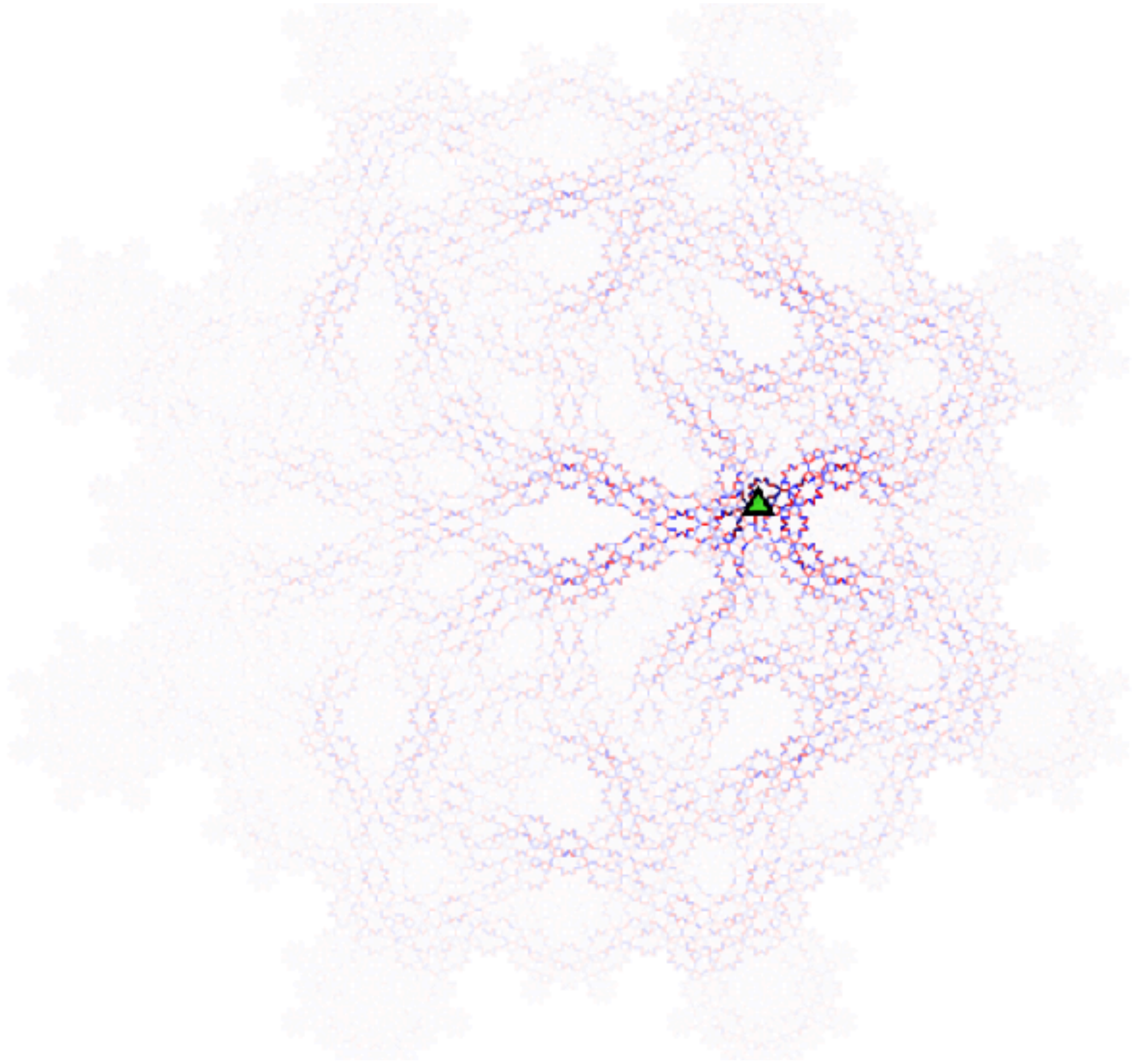
Criticality in Amman Beenaker (AB) dimer covers

Amman Beenaker Tiling



With dimers

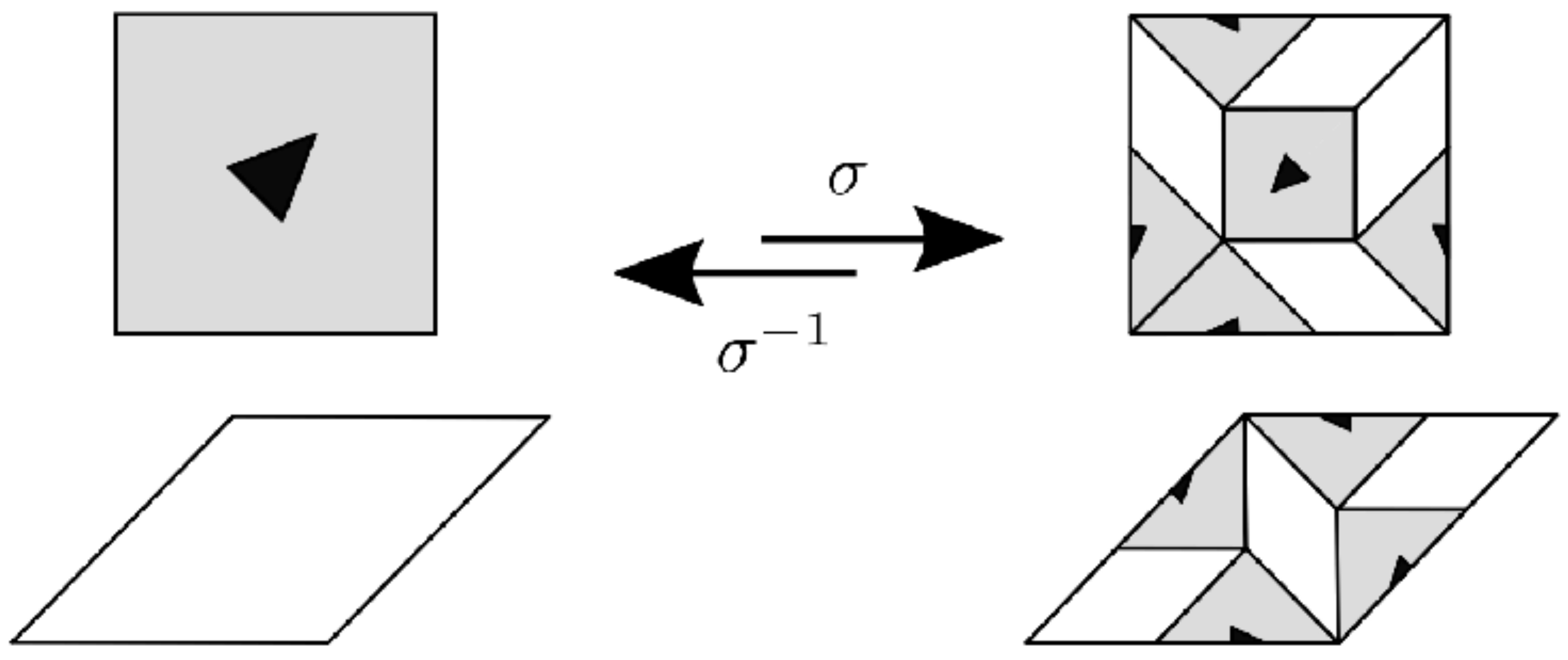


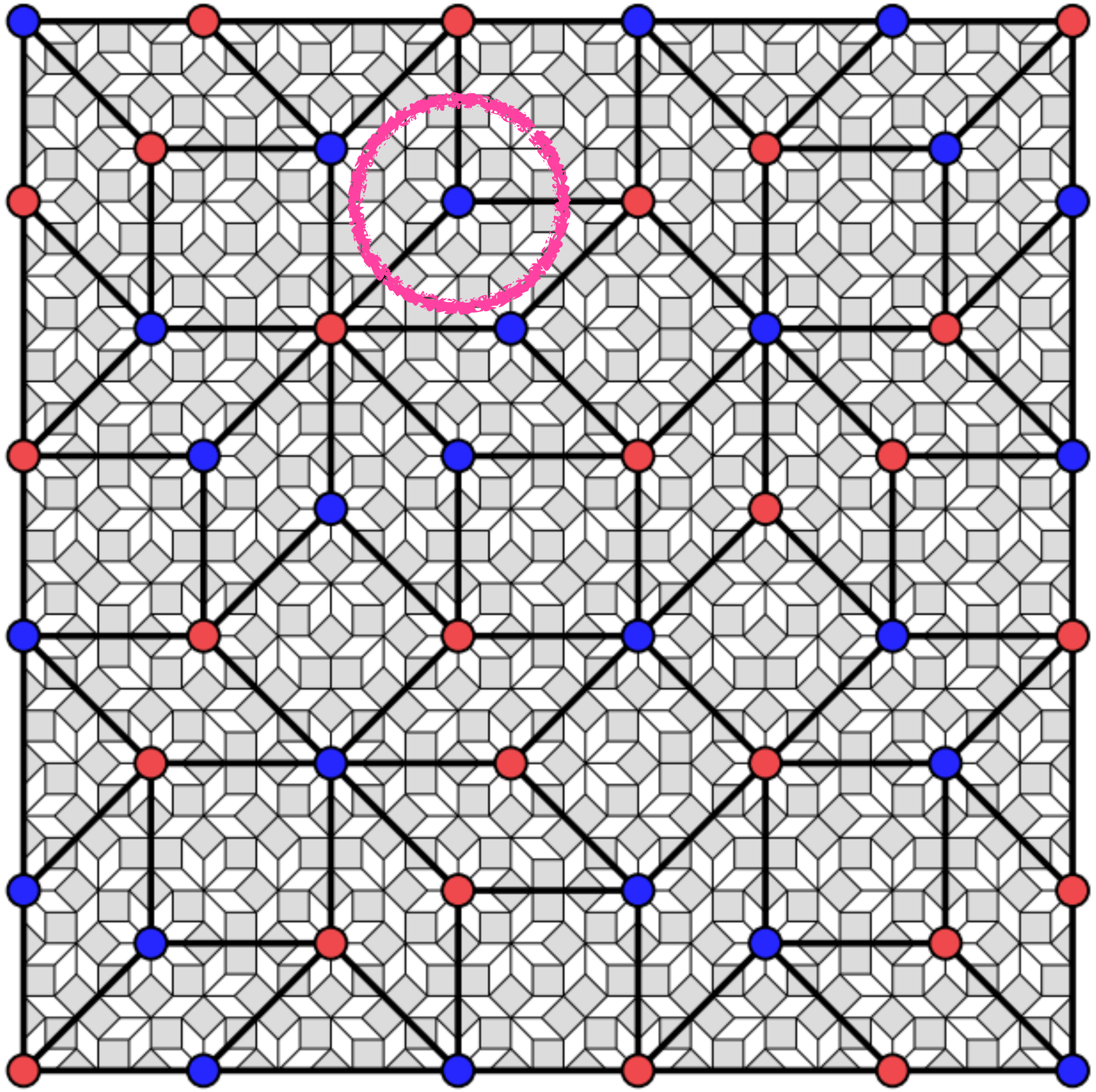


Statistical mechanics of dimers on quasiperiodic tilings

Jerome Lloyd,[1,2,3,*] Sounak Biswas,[1,*] Steven H. Simon,[1] S. A. Parameswaran,[1] and Felix Flicker[1,4]

Scale invariance is inherent to Quasi-crystals





V area
~ Coarse Graining Cell

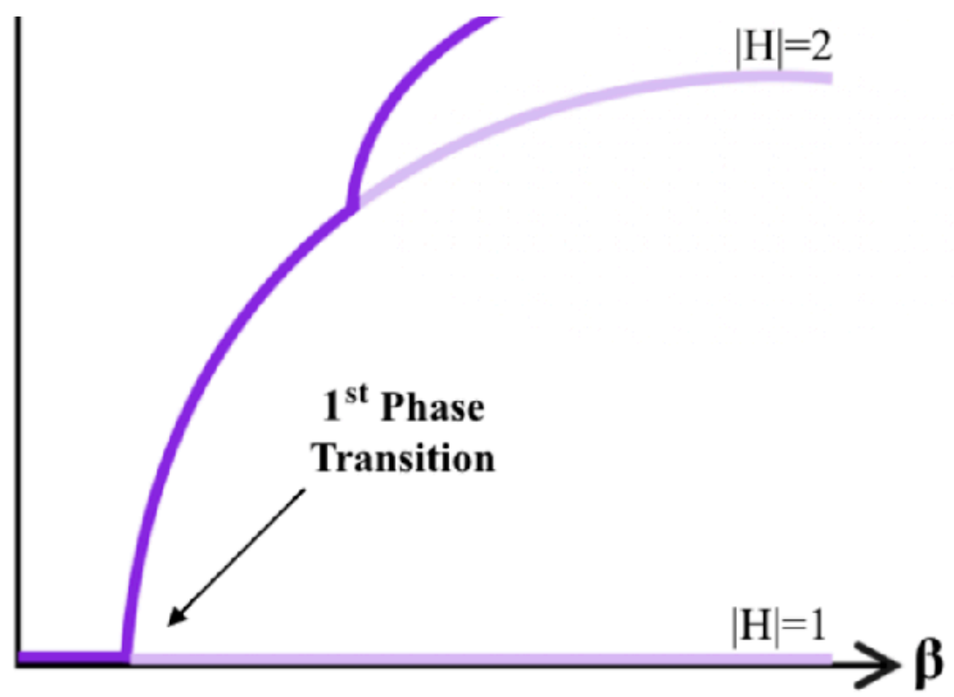
8-vertices
A sublattice

8-vertices
B sublattice

Apply our approach (RSMI version) - raw outputs

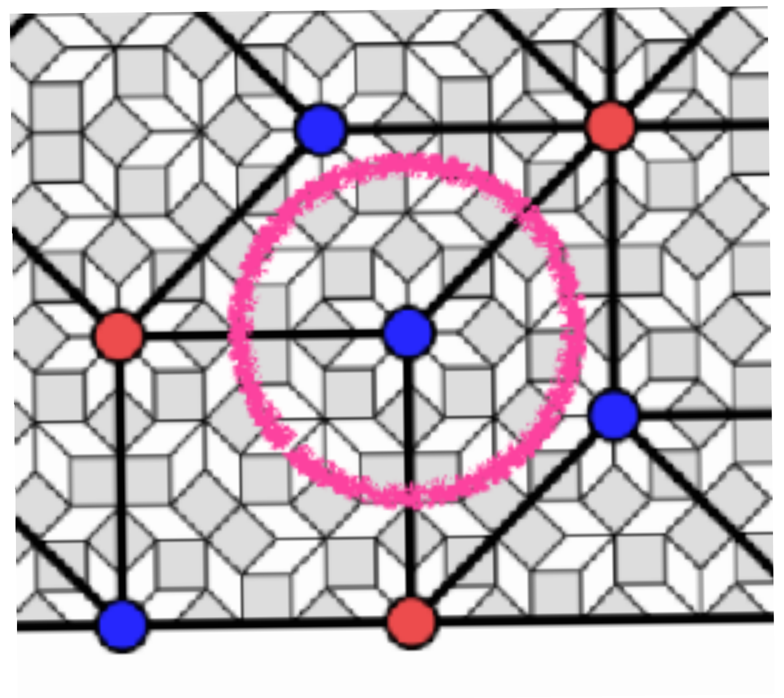
Setup

|H|=2 β=∞



Linear DNNs for P(H|V)

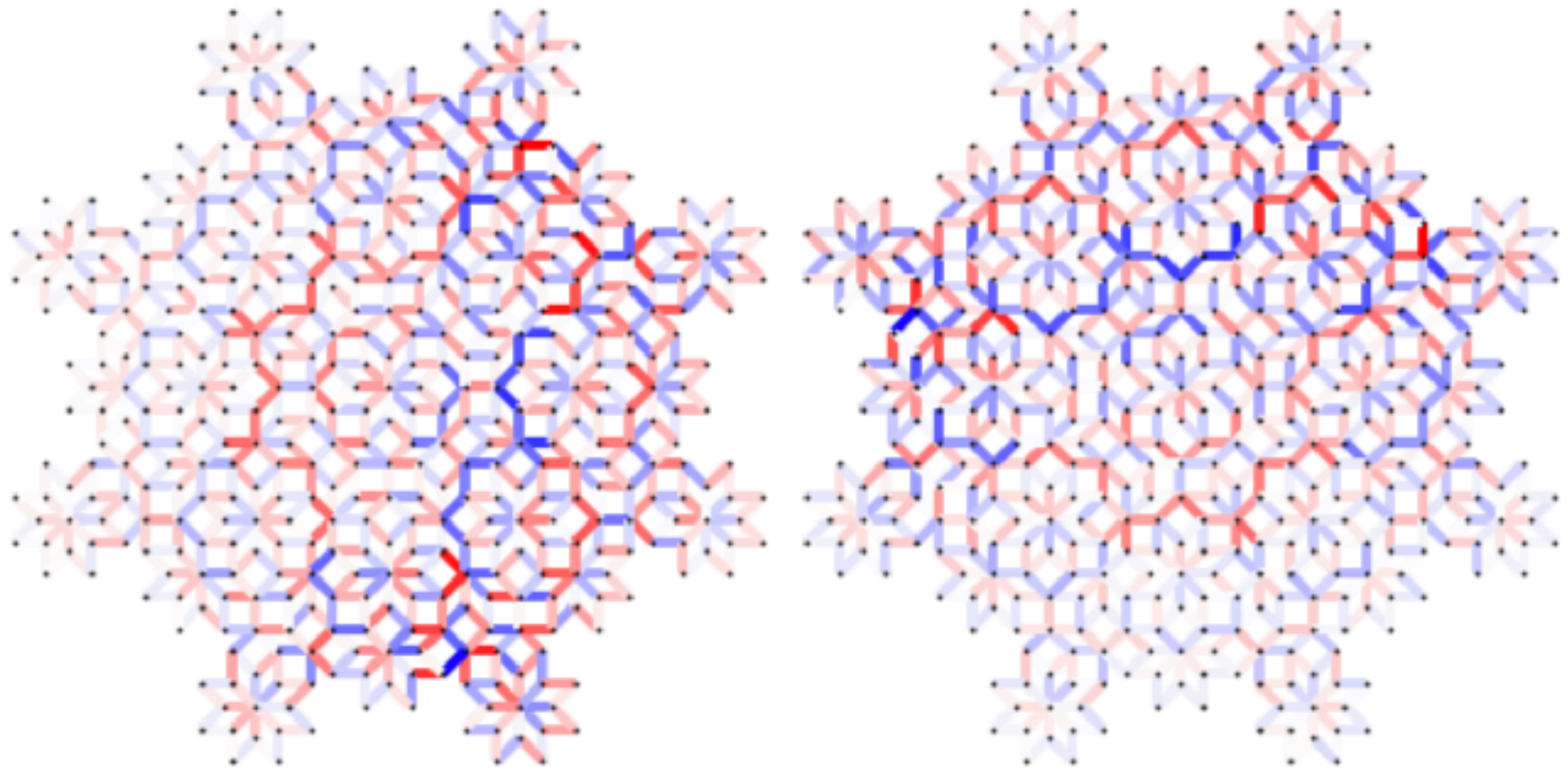
V=8-vertex which coarse grains to 3-vertex



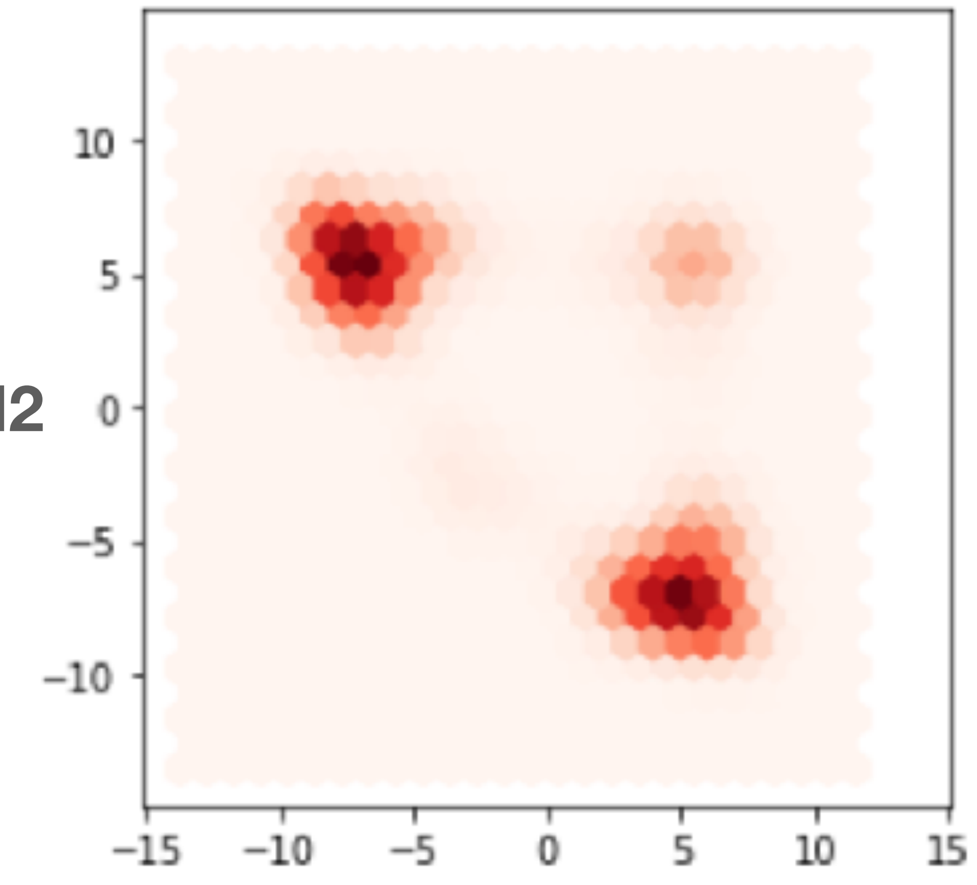
Results

H1

H2



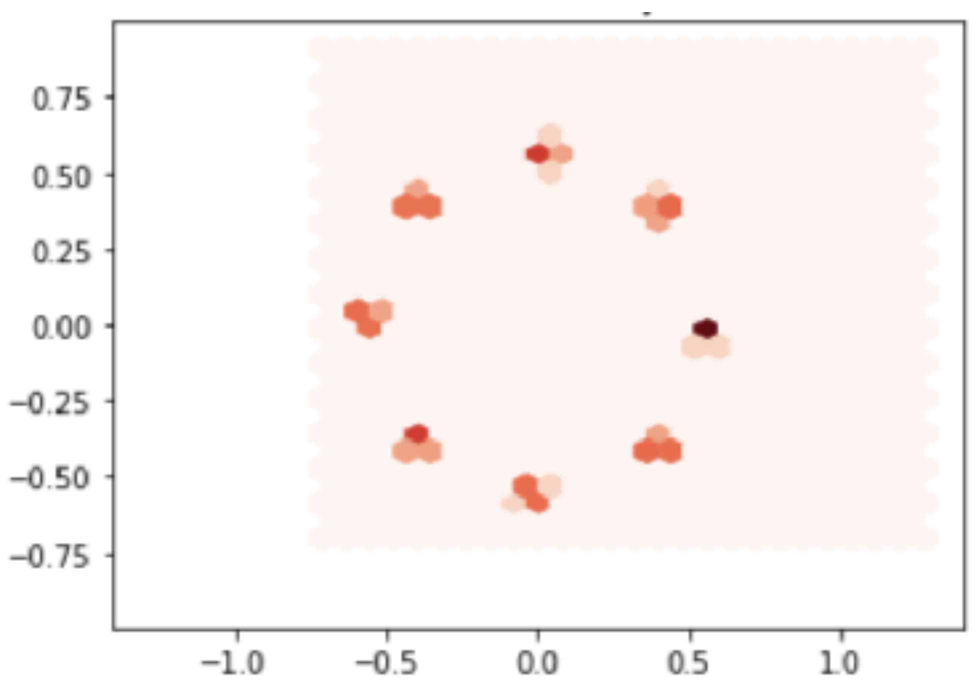
Histogram of H1 and H2

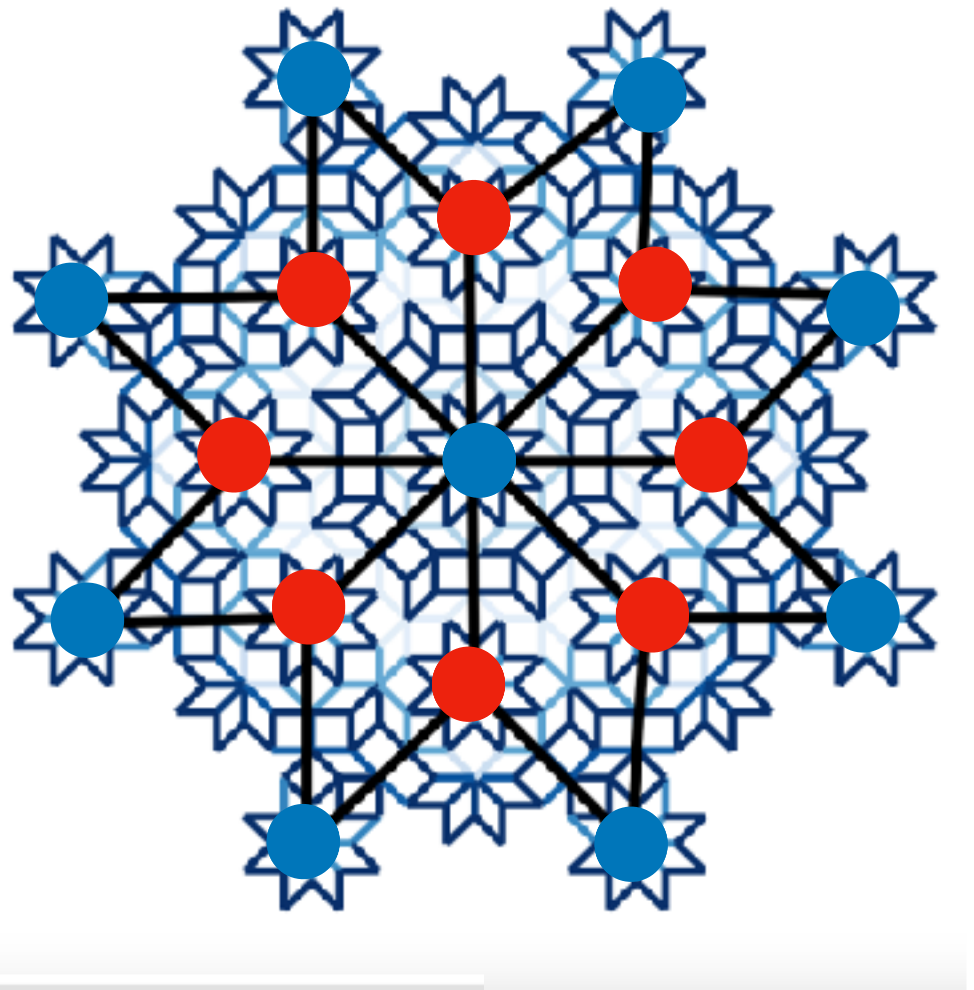


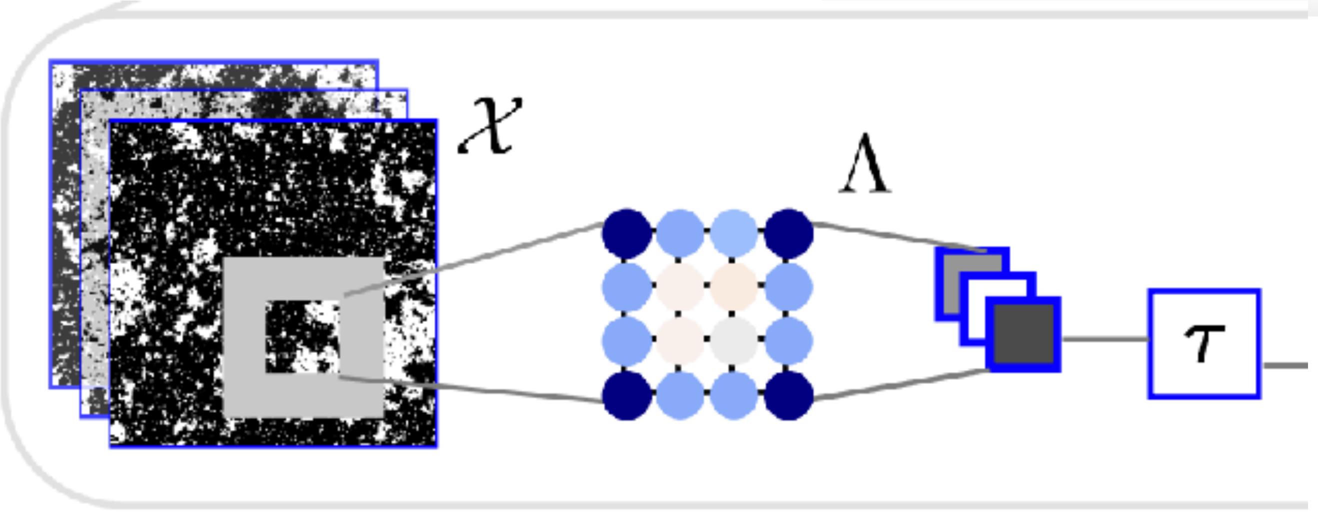
Biswas, Ringel, Flicker, Koch-Janusz - To be published

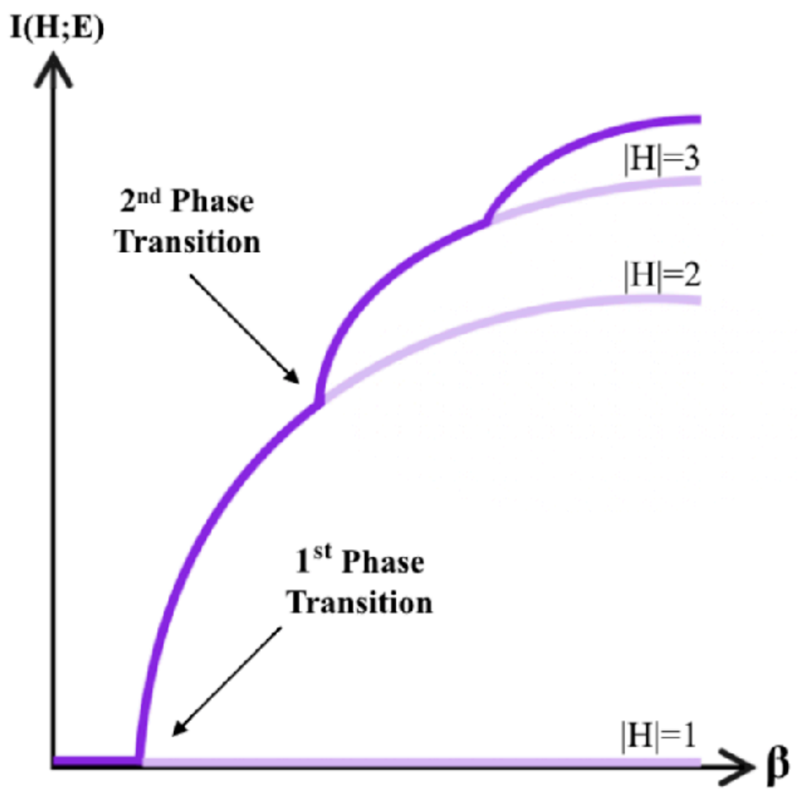
How to interpret these results?

- Find symmetries
- Examine P(H|V)
  - Visualize filters when based on a linear DNN
  - Find ideal input samples when based on generic DNN
- Scaling dimensions; Degeneracies; Representations
- Talk to the experts







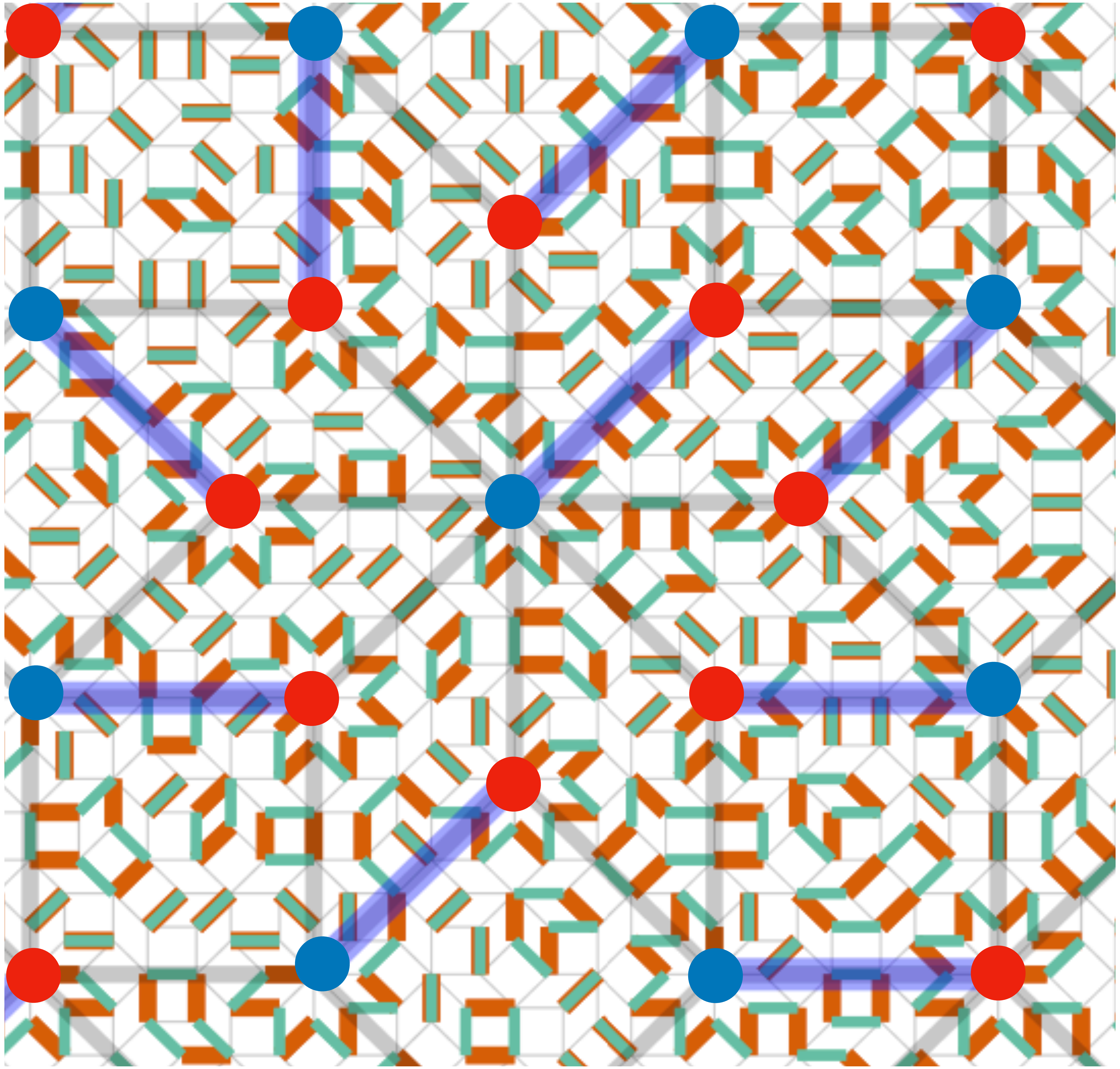


Dimer

Ref. Dimer

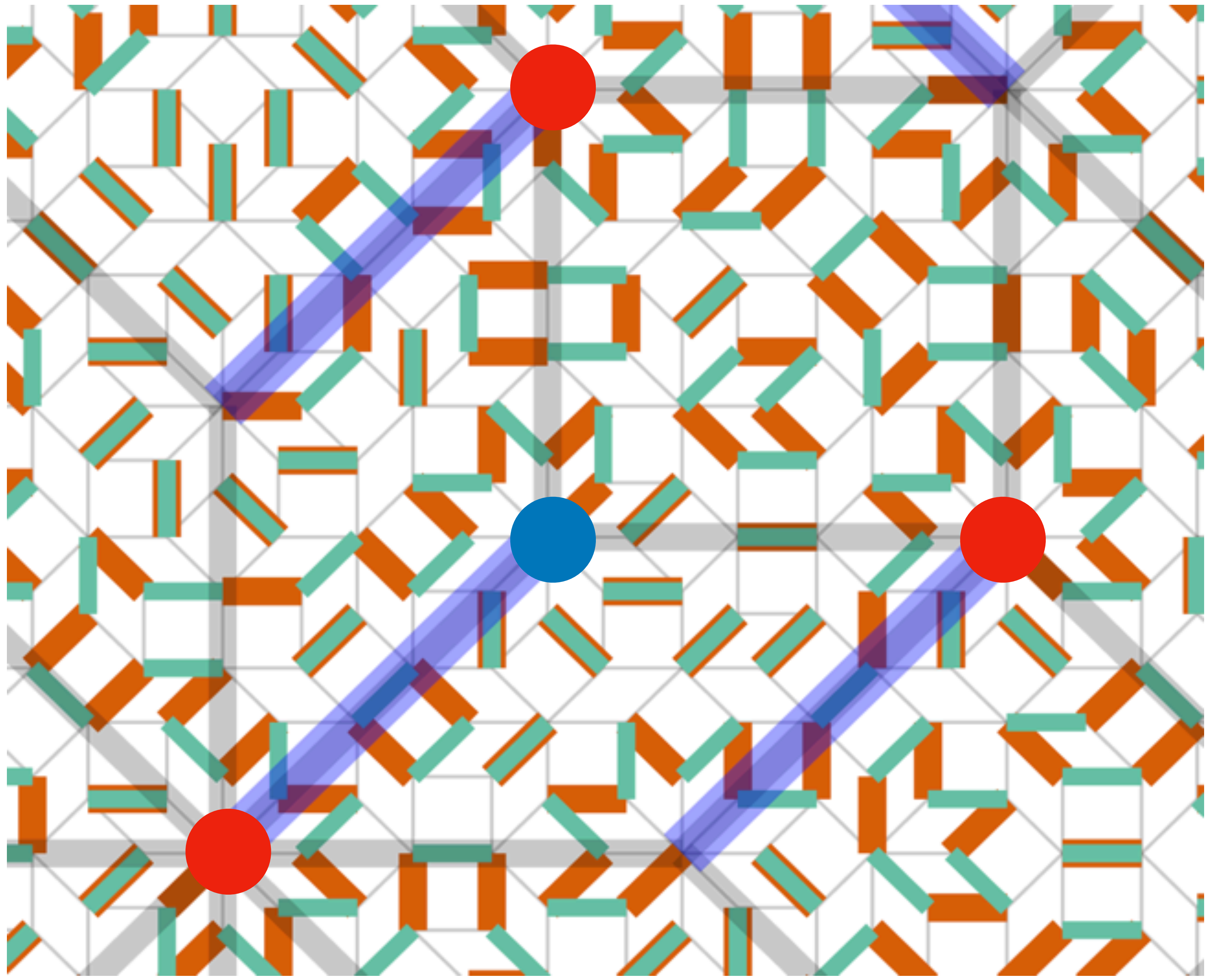
Coarse Grained Dimer

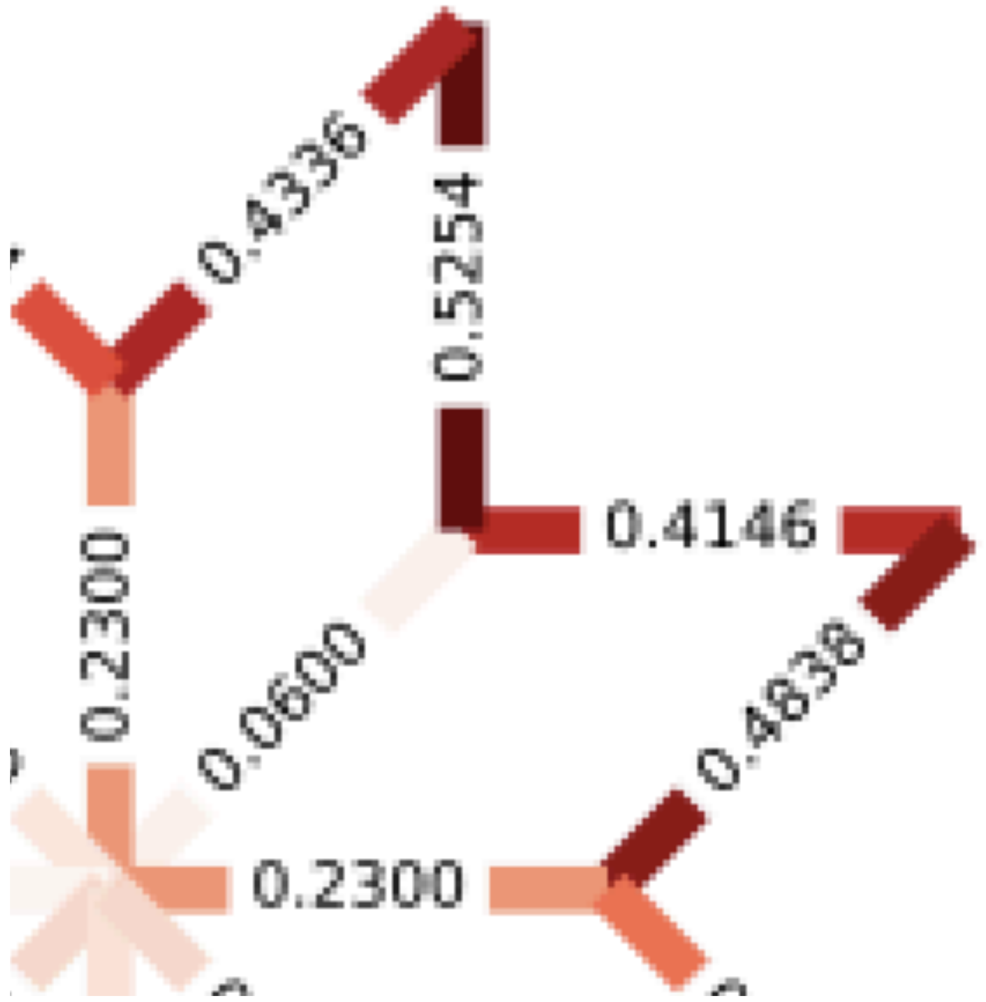
Alternating Path

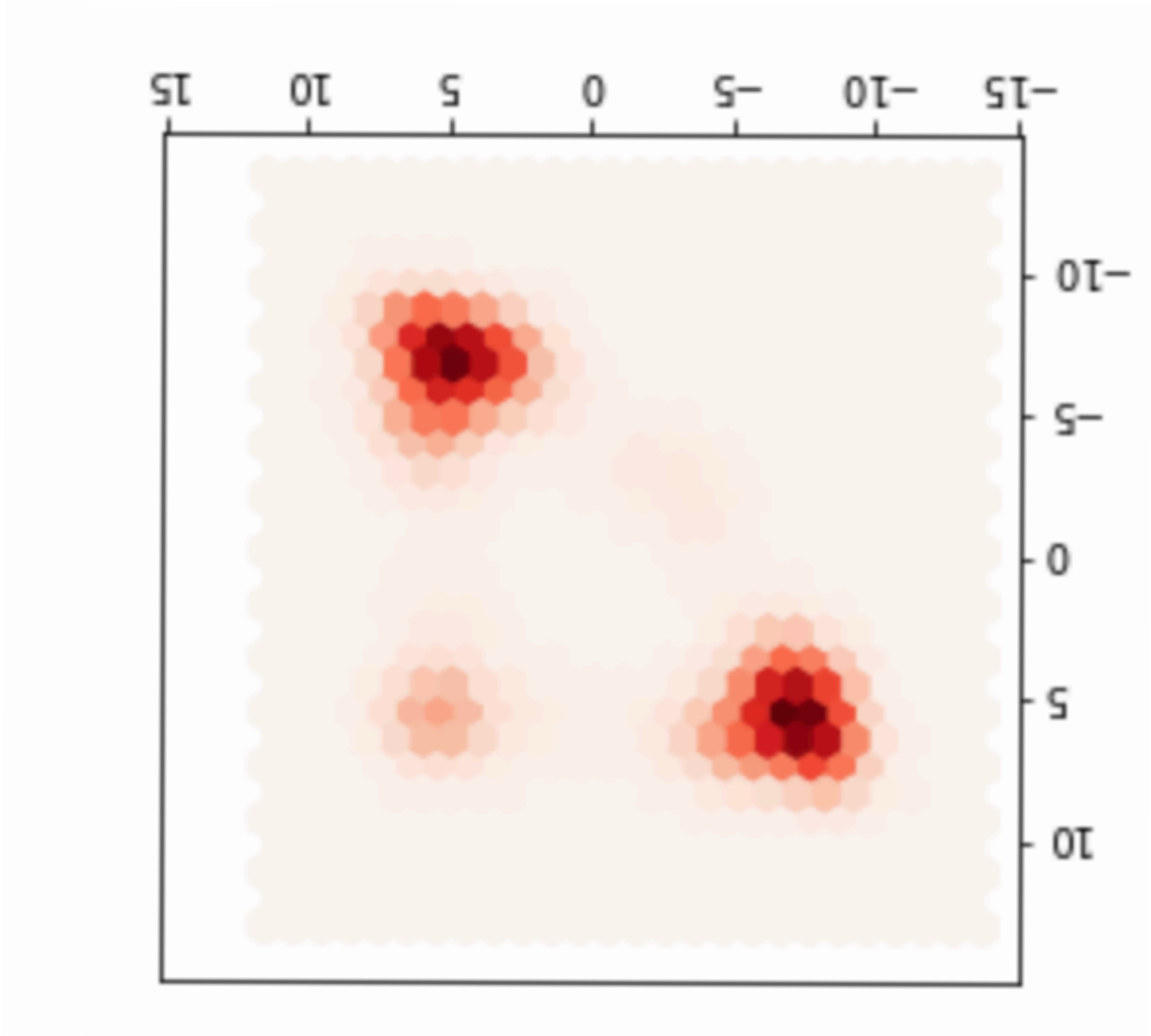


Biswas, Ringel, Flicker, Koch-Janusz - To be published
Also Biswas, Parameswaran - To be published

Conjecture - percolating paths are the coarse variables and behave as dimers

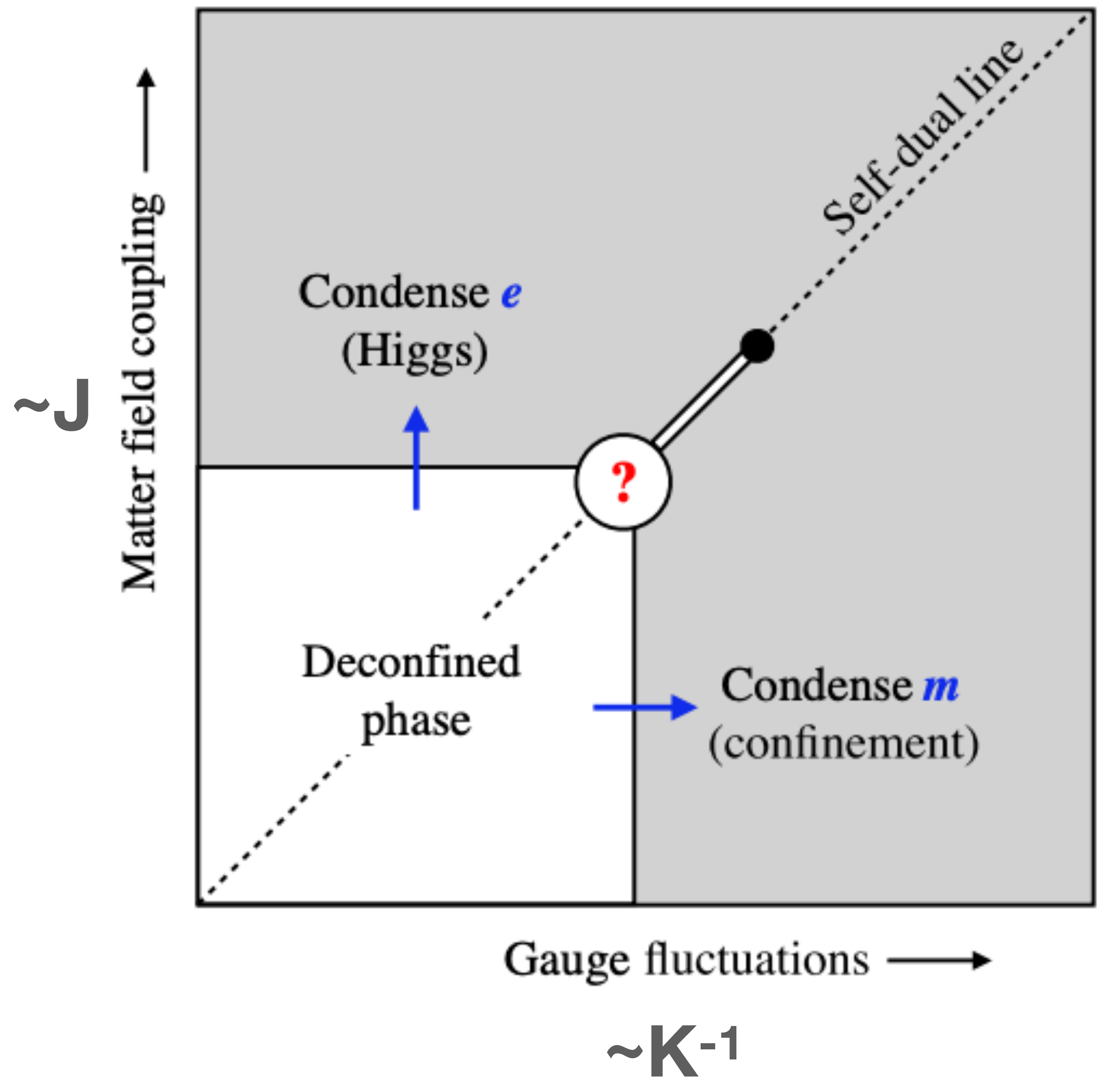




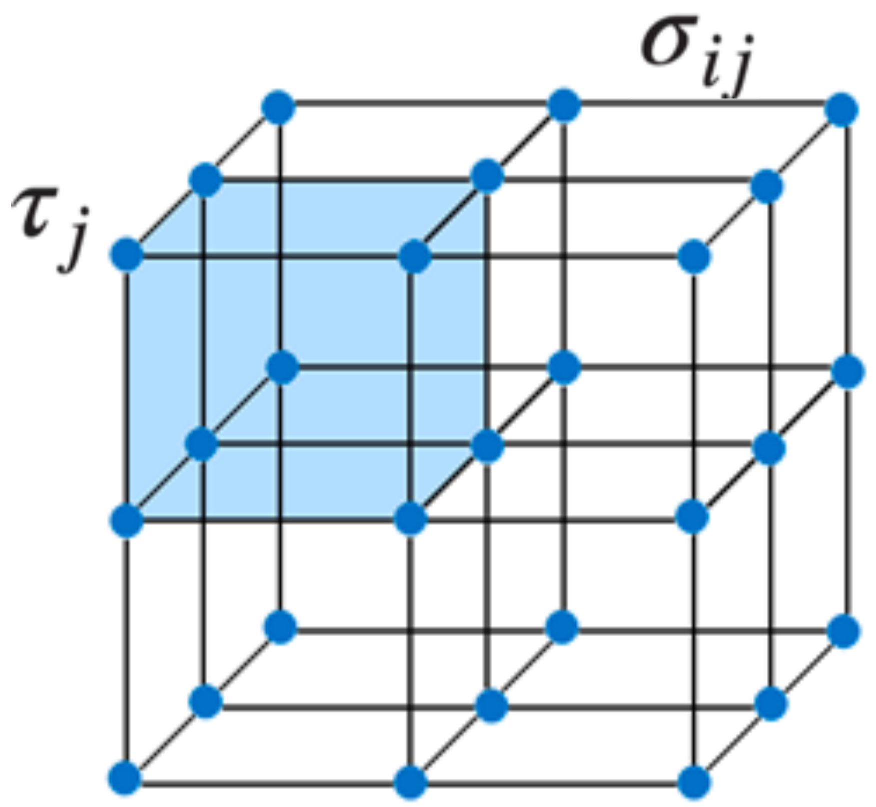


Biswas, Ringel, Flicker, Koch-Janusz - To be published

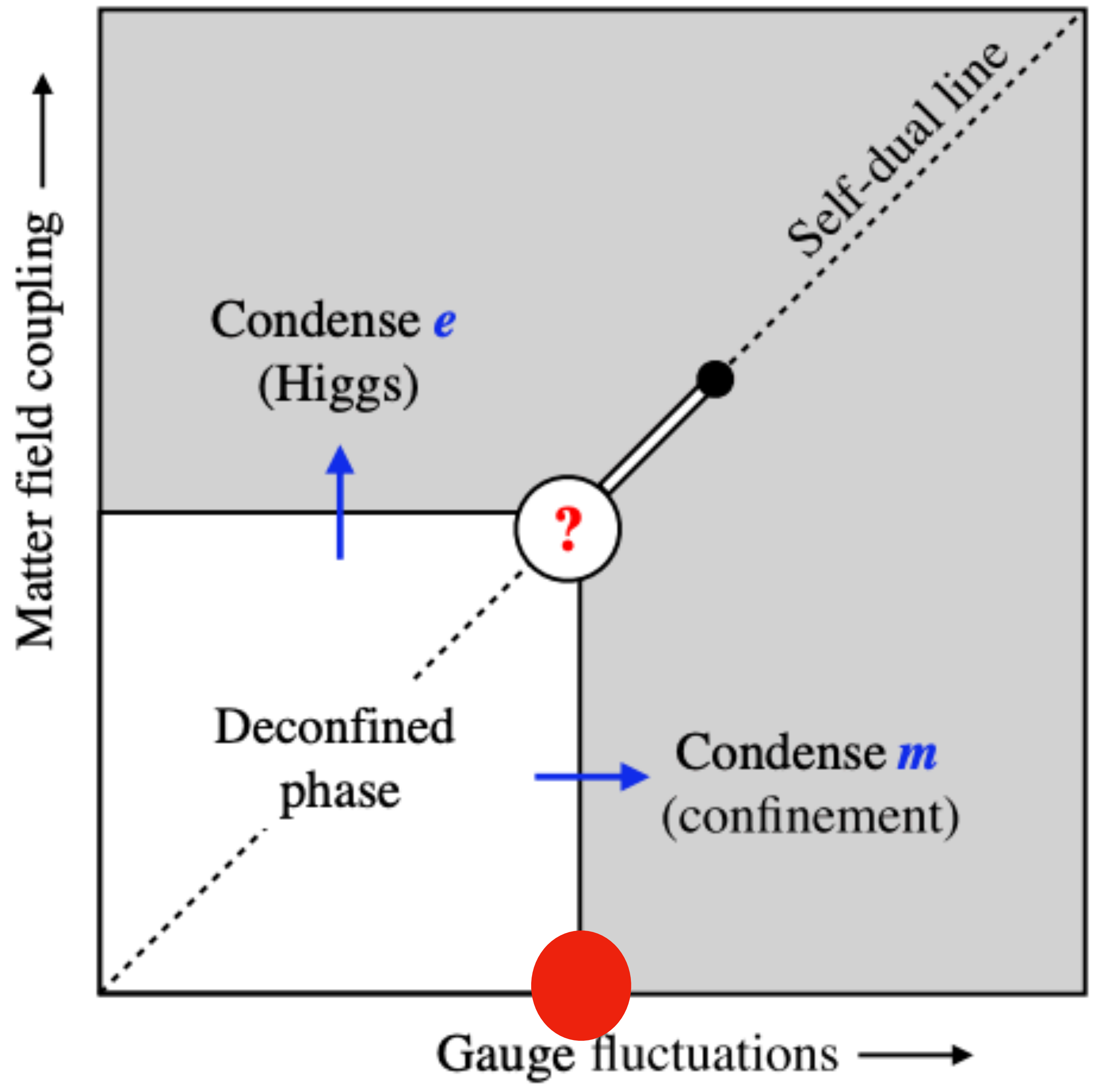
Self Dual Z2 Gauge-Higgs theory in 3d

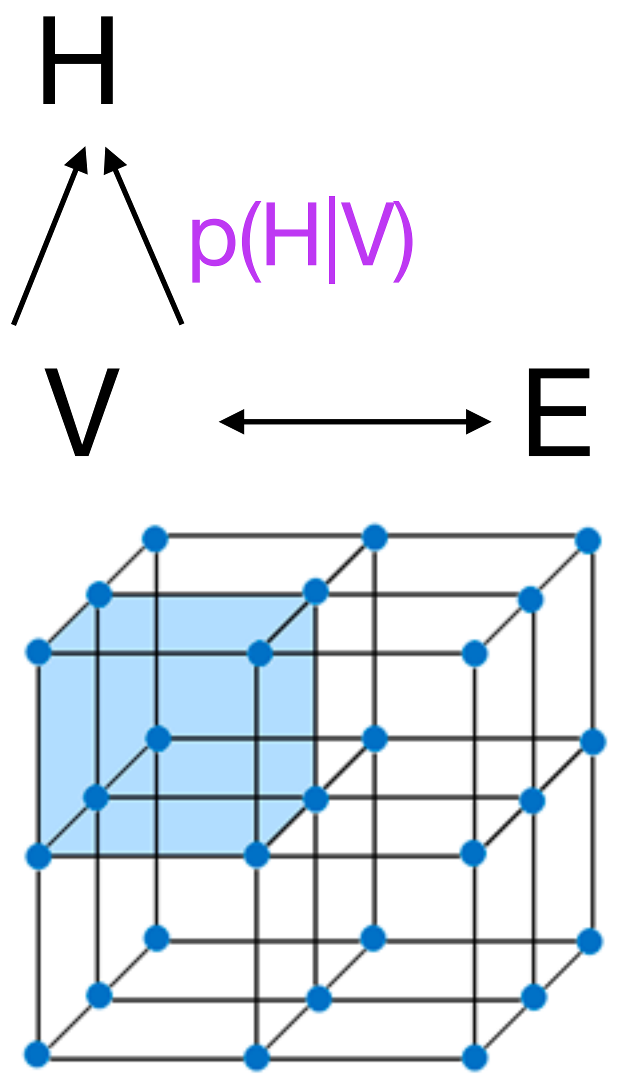


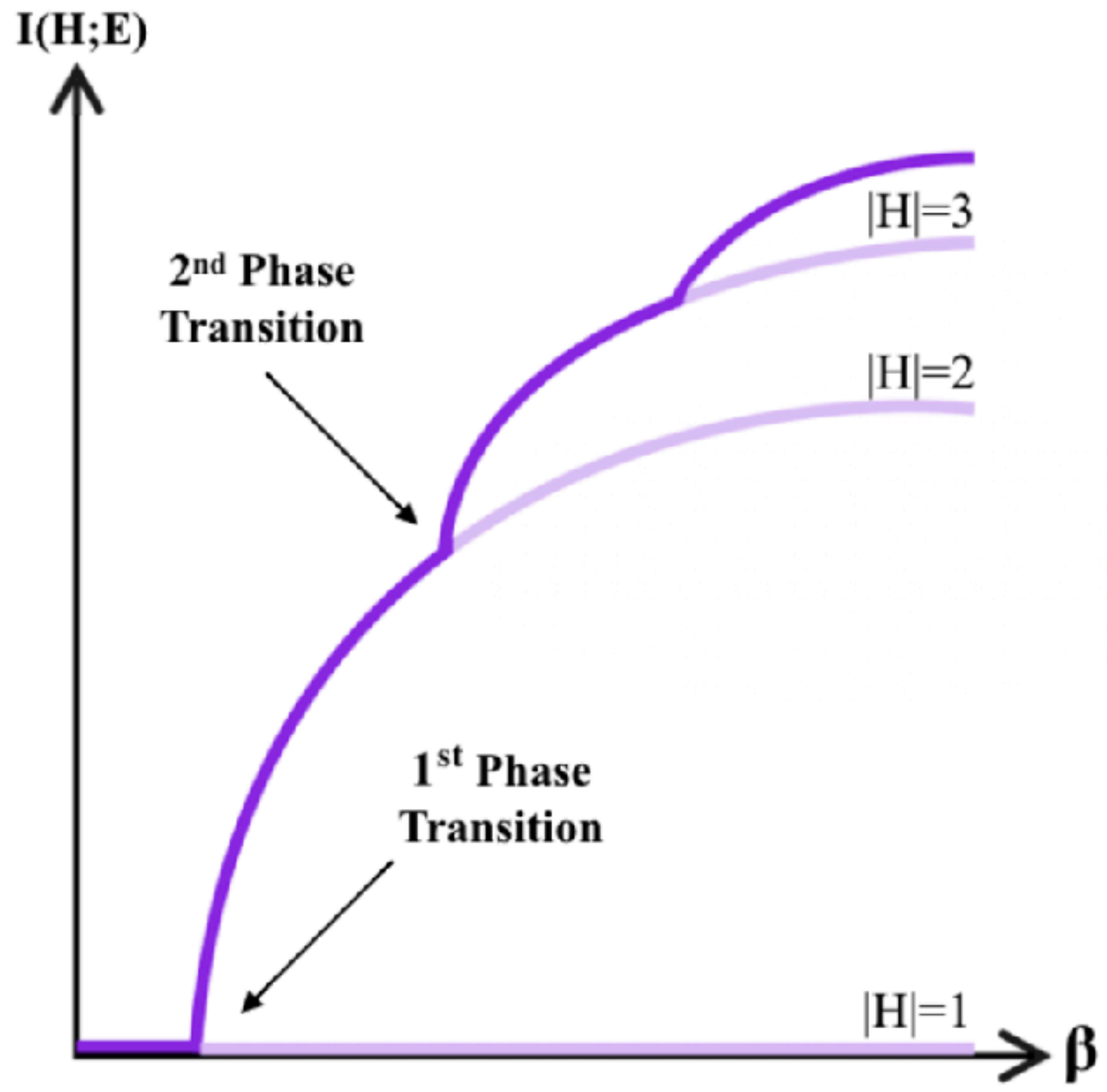
$$Z \propto \sum_{\{\sigma\},\{\tau\}} \exp\left(K \sum_{\square}\left(\prod_{\langle ij\rangle\in\square} \sigma_{ij}\right) + J\sum_{\langle ij\rangle} \tau_i \sigma_{ij} \tau_j\right)$$

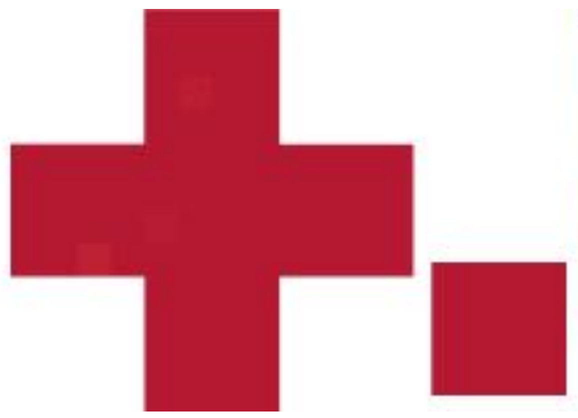


Apply our IB-RG approach - warm up





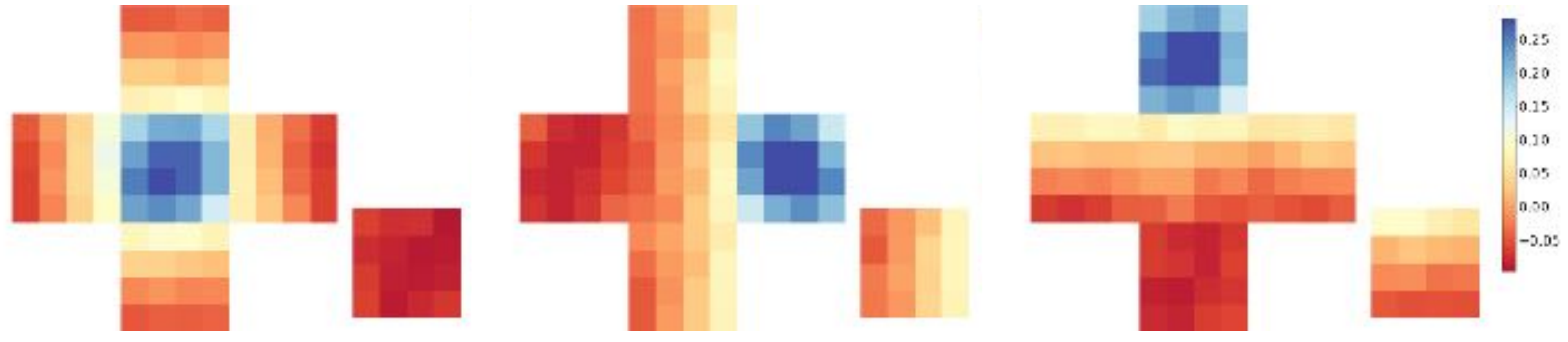




1st IB transition

Energy operator

Δ=1.41

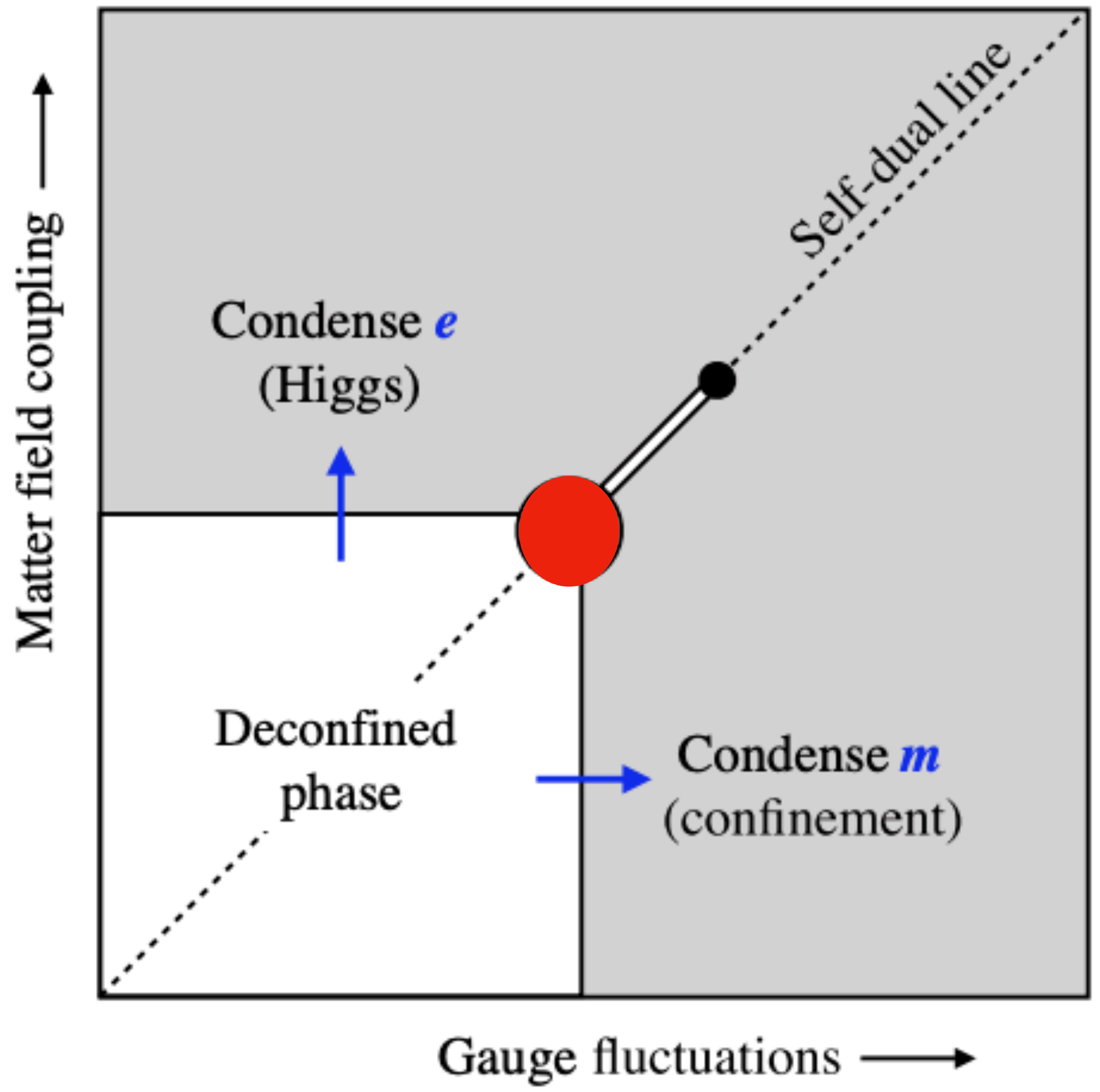


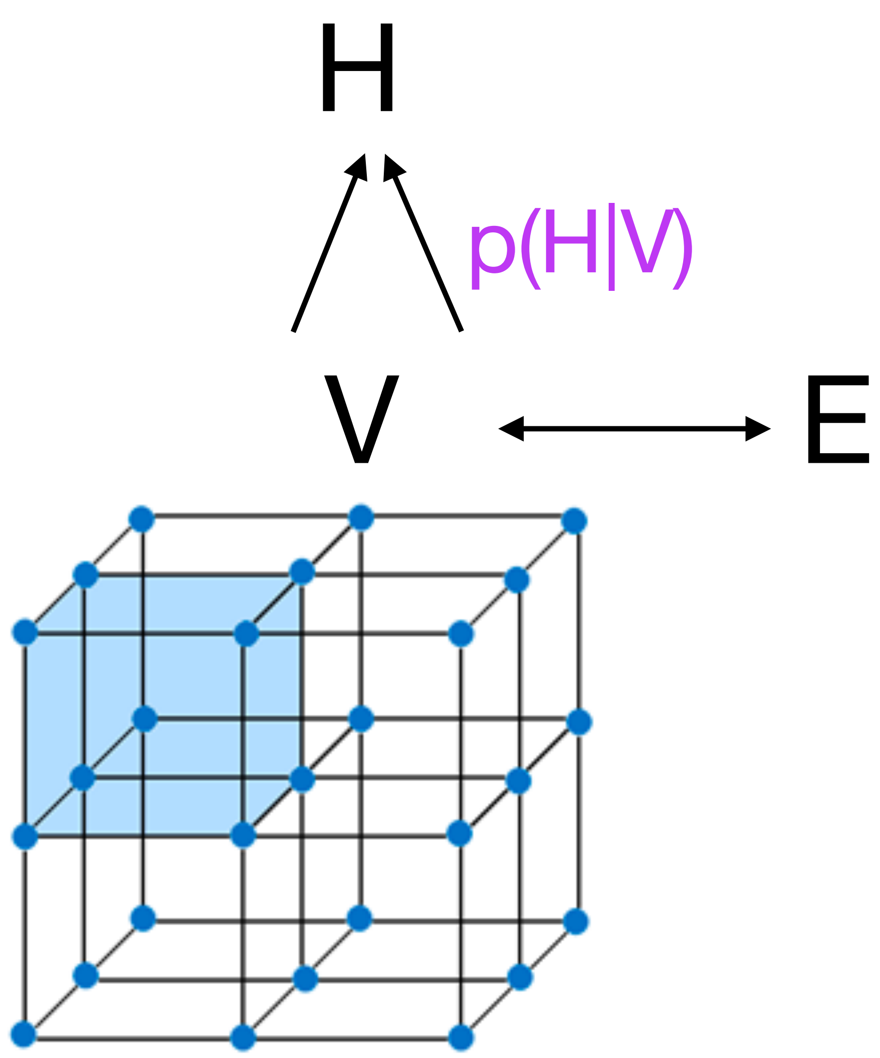
2nd IB transition

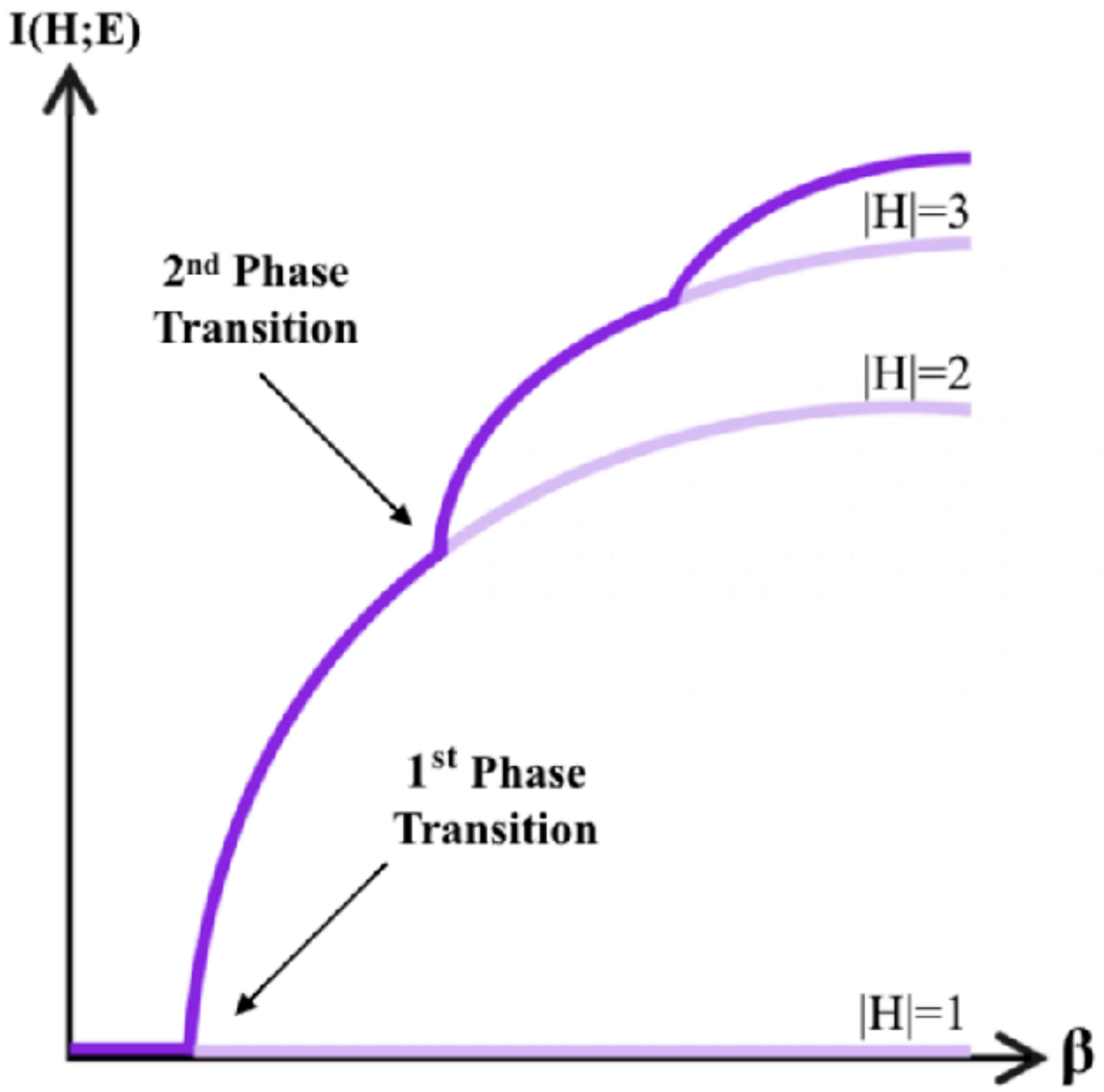
Three derivatives of energy operator

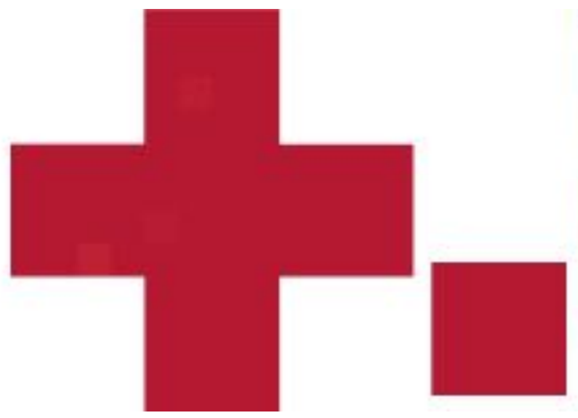
Δ=2.41

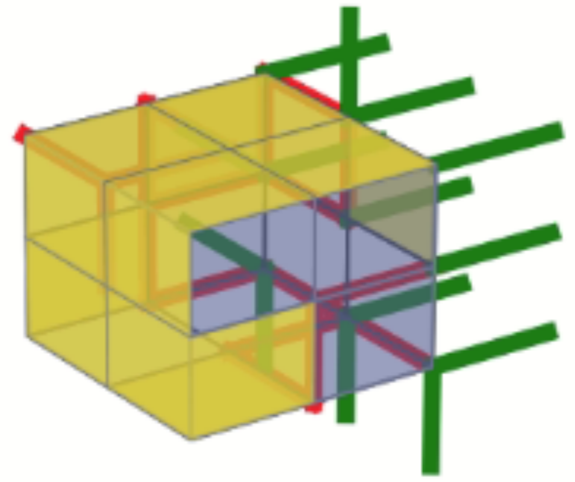
Apply our IB-RG approach - dual point











Emergent U(1) symmetry ?!

1st IB transition

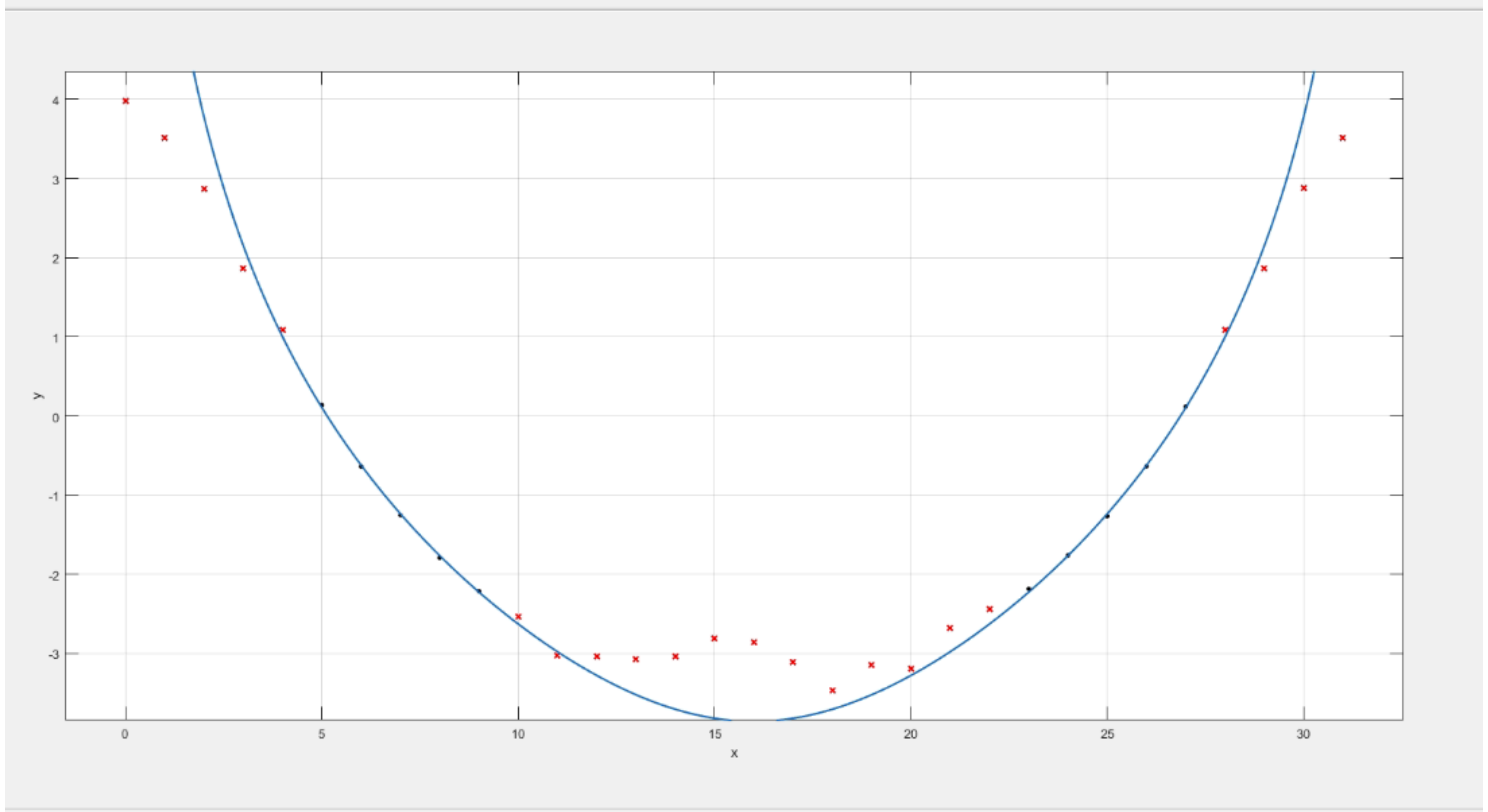
Energy operator

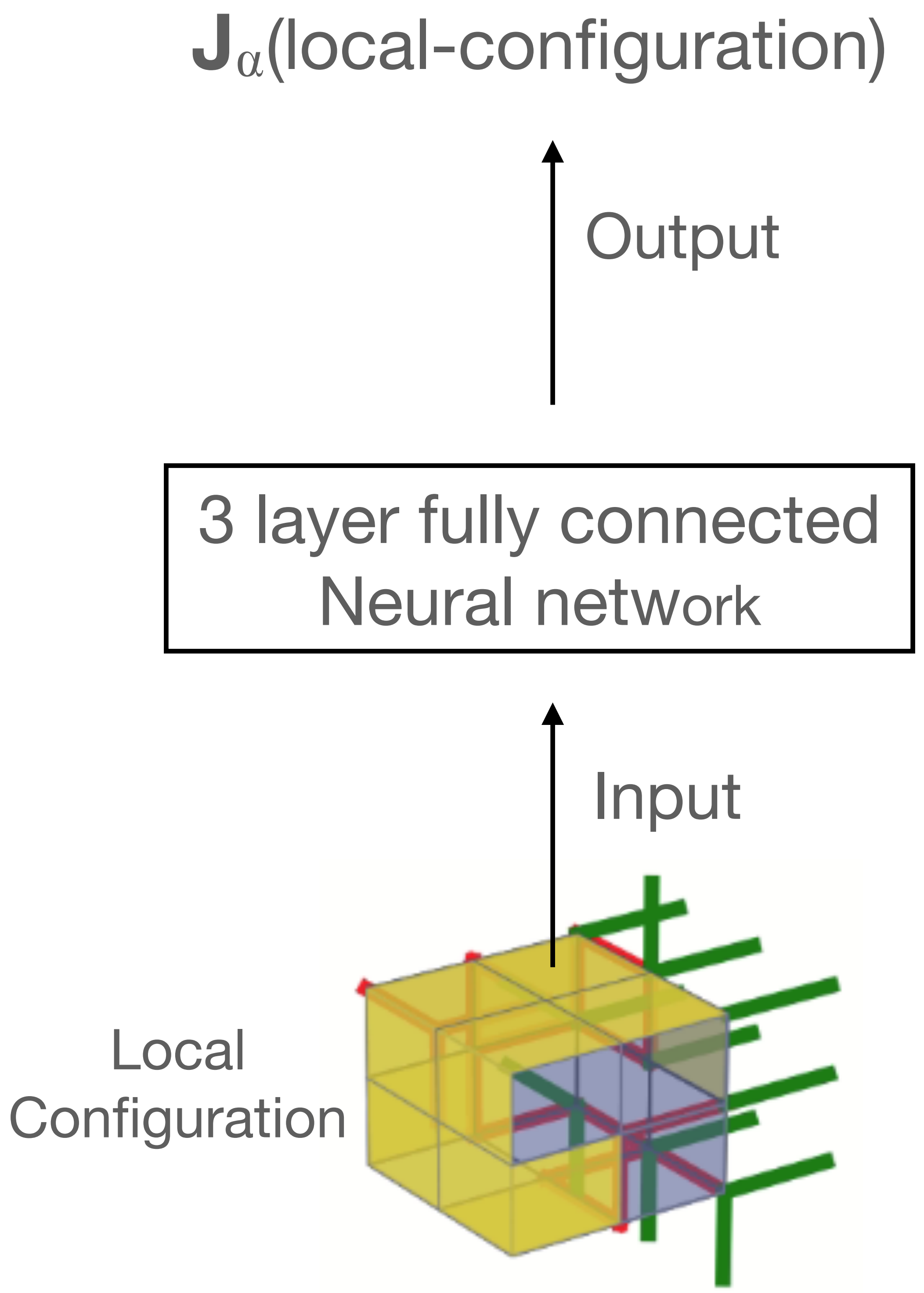
Δ=1.4-1.5

2nd IB transition

A non-scalar operator with scaling dimension 2

Suspected microscopic current operator (very preliminary…)

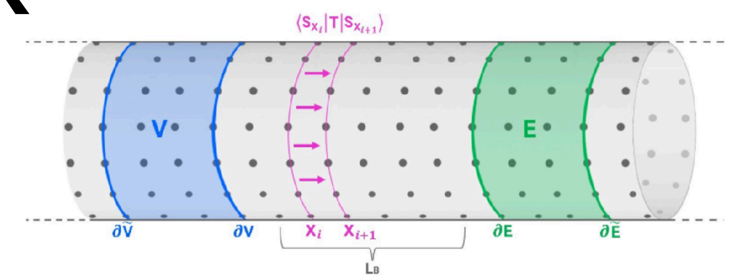




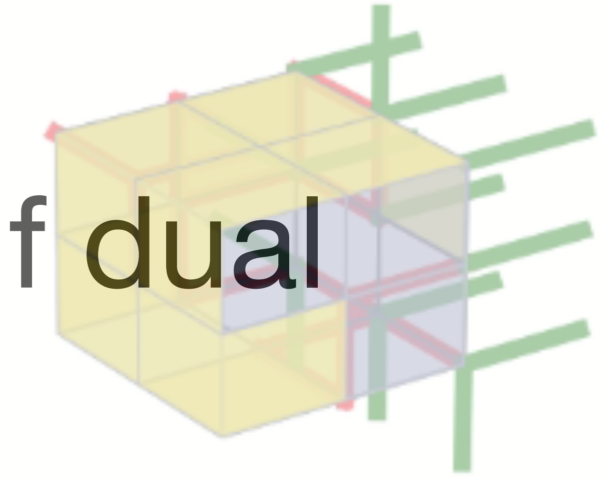
Oppenheim, Koch Janusz, Gazit, Ringel - To be published

Summary

- A concrete dictionary is emerging between information theory notion of relevance (IB) and physics notion of relevance (large-transfer-matrix eigenvectors/relevant-operators)



- Several deep learning based tools allow controlled numerically solutions of IB
- Applying this approach to two unsolved problems we get interesting preliminary results
    - Very preliminary evidence for an emergent current operator in a self dual gauge theory.



    - Note that the method also has the ability to refute the existence of such an operator.
    - We find a coarse graining method for a dimers on a quasi-periodi