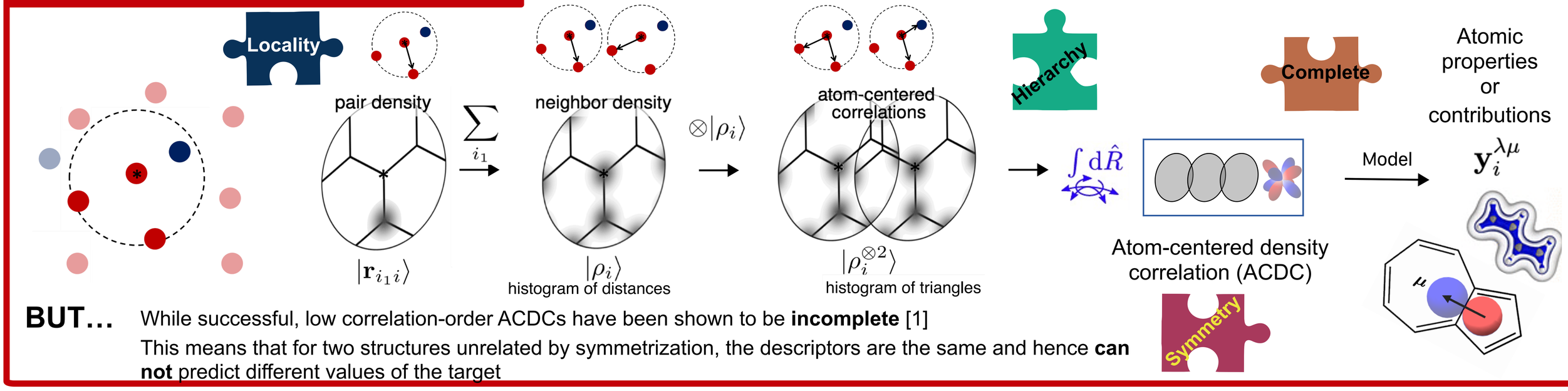
# Completeness of representations in atomistic machine learning

**Jigyasa Nigam,** Sergey Pozdnyakov, Kevin Huguenin and Michele Ceriotti

*Laboratory of Computational science and Modeling,*

*Institute des Matériaux, École Polytechnique Fédérale de Lausanne, Switzerland*

## ACDC's rock atomistic ML

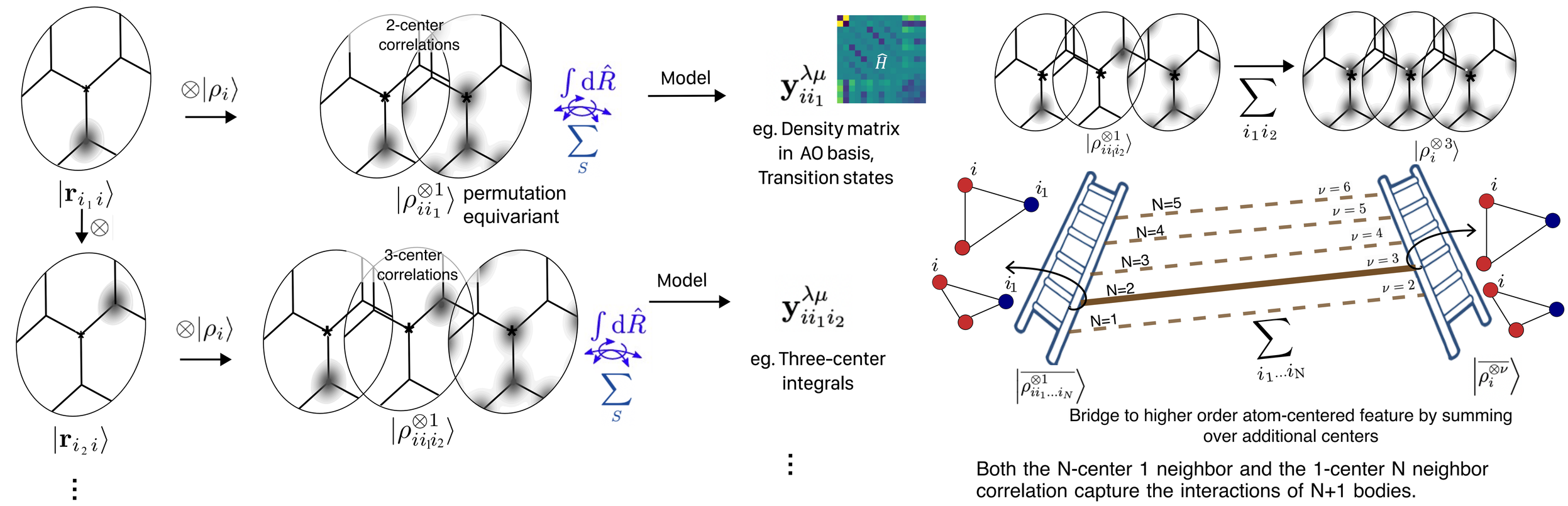



**BUT…** While successful, low correlation-order ACDCs have been shown to be **incomplete** [1]

This means that for two structures unrelated by symmetrization, the descriptors are the same and hence **can not** predict different values of the target

## Bricks to bridges

We have the ingredient to tag individual neighbors with respect to the center (pair density) — together with the ACDC, this forms a description of multiple centers, that can be then rotationally symmetrized the same way as above. [4]

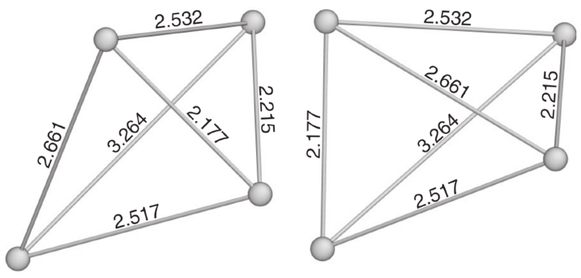


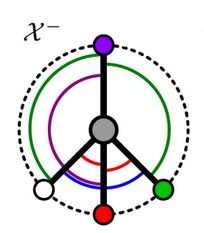

Bridge to higher order atom-centered feature by summing over additional centers

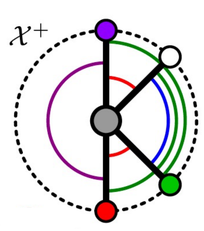
Both the N-center 1 neighbor and the 1-center N neighbor correlation capture the interactions of N+1 bodies.

In addition to modeling properties indexed by more than one center, higher-center representations may be used in the intermediate modeling steps, e.g. 2-center contributions. [3,5]

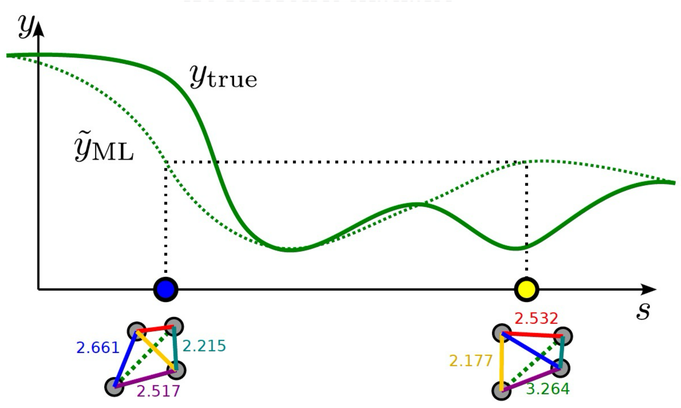
## Houston, we have a problem!



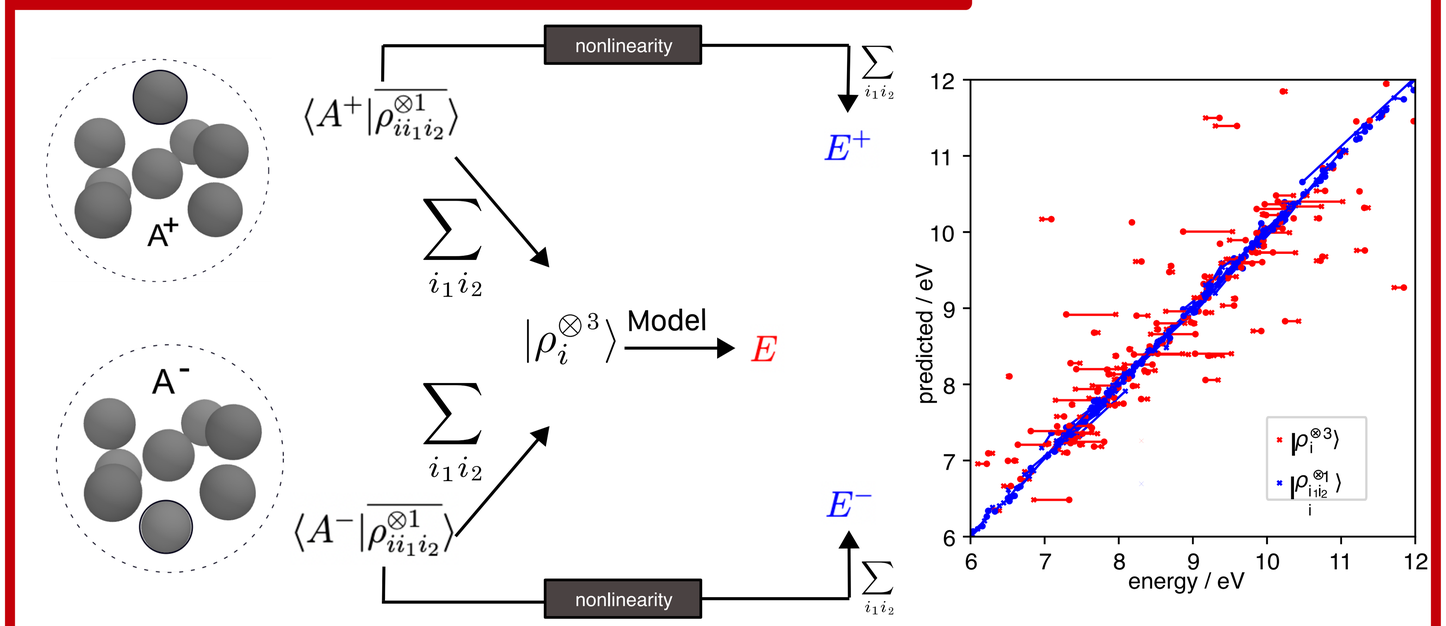


Examples of pair of structures that have the same set of pairwise distances [2] thus proving 2-body correlations are incomplete. The numerical instability caused by this degeneracy of descriptors is visualized below.



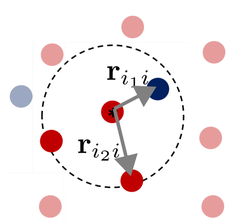

Ref. 1 proposed manifolds of structures that have the same set of pairwise distances and angles thus proving that 3 (and 4)-body correlations are also incomplete.

## But, every problem has a solution




Pair of structures (unrelated by symmetry)[1] and having different energies) that have the same bispectrum. Therefore a model built on bispectrum can not distinguish the energies of the two structures (line segments joining the predictions of the pair energies )

$\mathbf{r}_{i_1 i}$, $\mathbf{r}_{i_2 i}$ $\mathbf{r}_{i_1 i} \times \mathbf{r}_{i_2 i}$ forms the unique right-handed coordinate system to express neighbor positions. The three center- one neighbor correlation is the histogram of distances relative to this frame and can hence retains complete information about the structure. [6]



## It's a wrap (for now)

- Thus N-centered representations serve as the playground for understanding the relationship between the representation (the architecture) and the physical target being modeled and mitigates deficiencies in existing schemes.
- ***Dream of communication***: Having information on atomic environments and their connectivities (graph N-centers) we have the basic ingredients for message-passing (MP). Can these structural representations be a bridge to understanding MP based neural networks? [7,8]

## Read on

1. S. Pozdnyakov et al. Phys. Rev. Lett. 125 (2020)
2. A.P. Bartok et al. Physical. Rev. B 87 (2013)
3. L. Zhang et al. npj Comp. Mat. 8(1) (2022)
4. J. Nigam et al. J. Chem Phys 156, (2022)
5. A. Musaelian, S. Batzner, et al. Nat. Comm.18(2023)
6. J. Nigam et al. APL Machine Learning 2, (2024)
7. J. Nigam et al. J. Chem Phys 156, (2022)
8. I. Batatia et al. arXiv 2205.06643 (2022)

## More at

*jigyasa.nigam@epfl.ch*

curiosity54
lab_cosmo