

Lecture 2. ML for Spectroscopy and Imaging Data

Sergei V. Kalinin

University of Tennessee, Knoxville and
Pacific Northwest National Laboratory



Spectroscopic Imaging

Scanning probe microscopy:

- Force-distance curve measurements
- Current-voltage measurements
- Piezoresponse force/electrochemical strain spectroscopy

Electron microscopy:

- Electron Energy Loss Spectroscopy

Optical microscopy:

- Hyperspectral imaging
- Time resolved measurements

Mass-spectrometry:

- Secondary ion MS imaging

In many cases, measured signal can be represented or approximated as a linear combination of signals. However, their functional forms are generally unknown

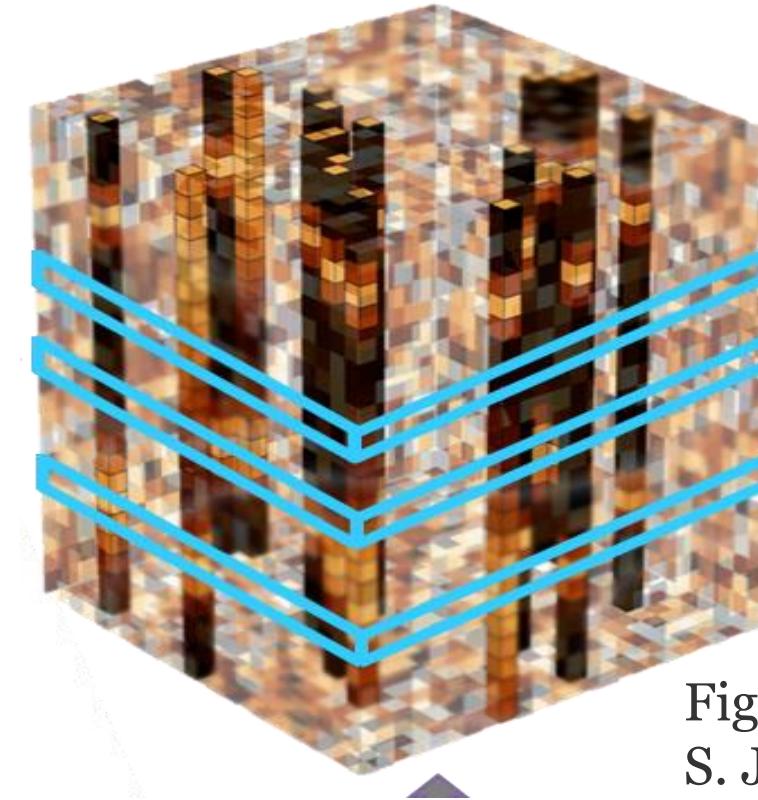
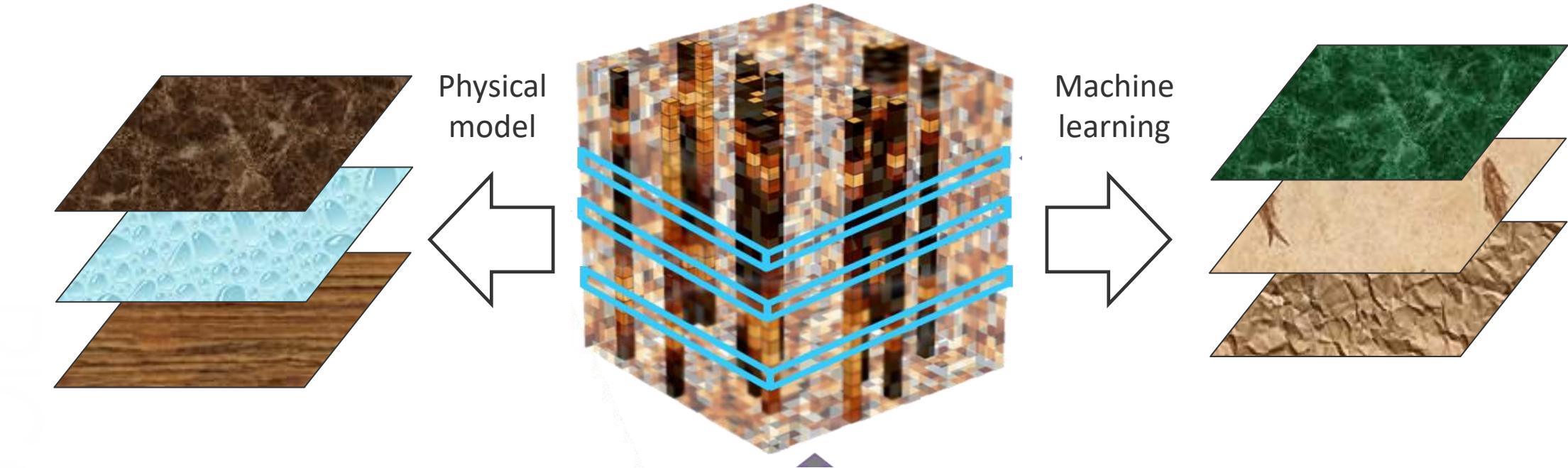


Figure by
S. Jesse

Very important: convolution with resolution function is also mixing

Physics vs. ML based analysis



- If we have physical model, we can extract relevant parameters from data
- Imperfect model: epistemic uncertainty
- Noisy data: aleatoric uncertainty
- Analysis results do not depend on sampling of data in x,y
- If we don't have physical model, we can learn intrinsic structure of data
- **Unsupervised learning:** based on data only
 - But not really (definition of distance)
 - Analysis can depend on sampling of data
- **Supervised learning:** based on prior examples
 - Out of distribution shifts

Physics-informed ML: Combines strengths (and limitations) of both approaches

General linear unmixing

$$S(\mathbf{x}, \mathbf{R}) = \sum_i a_i(\mathbf{x}) w_i(\mathbf{R}) + N$$

We start with:

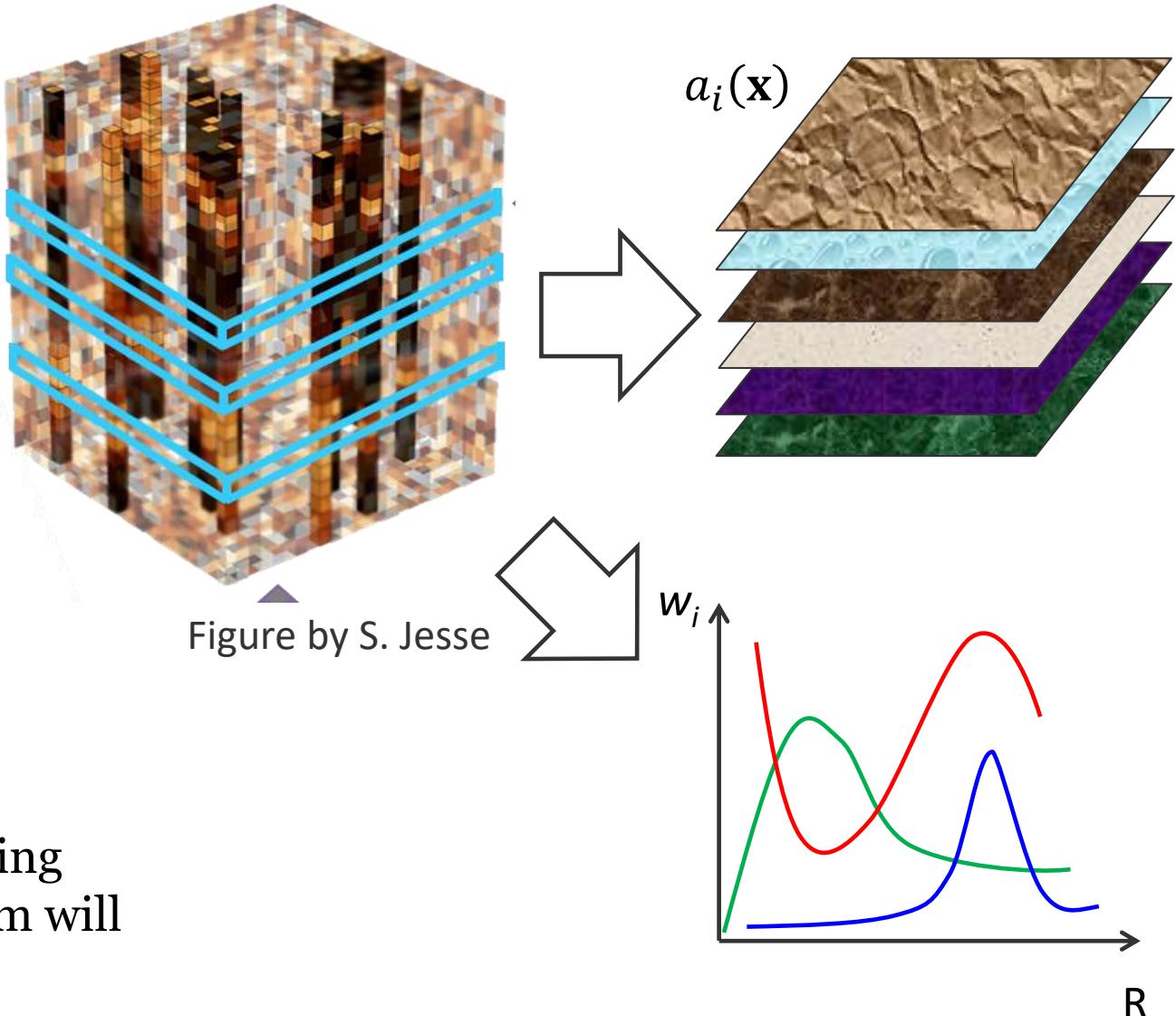
- \mathbf{x} is the spatial variable, $\mathbf{x} = (x, y)$
- \mathbf{R} is the (vector) parameter variable

Overall, for $M \times M$ image and P point in spectra, we have $M^2 P$ data points

We aim to get:

- $a_i(\mathbf{x})$ are loading maps
- $w_i(\mathbf{R})$ are endmembers/eigenvectors
- N is noise

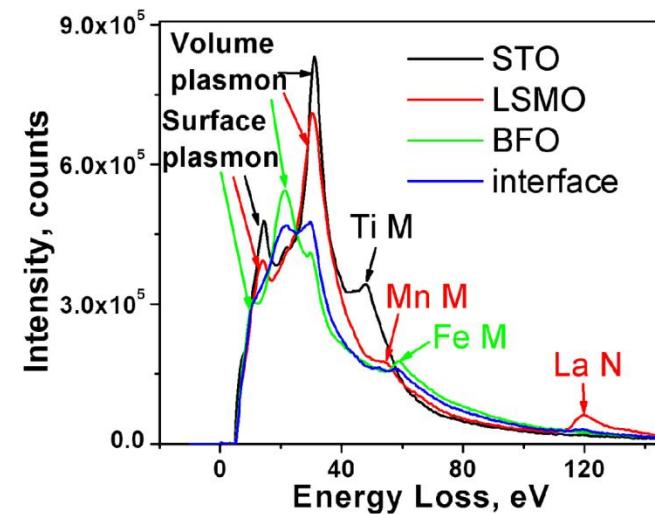
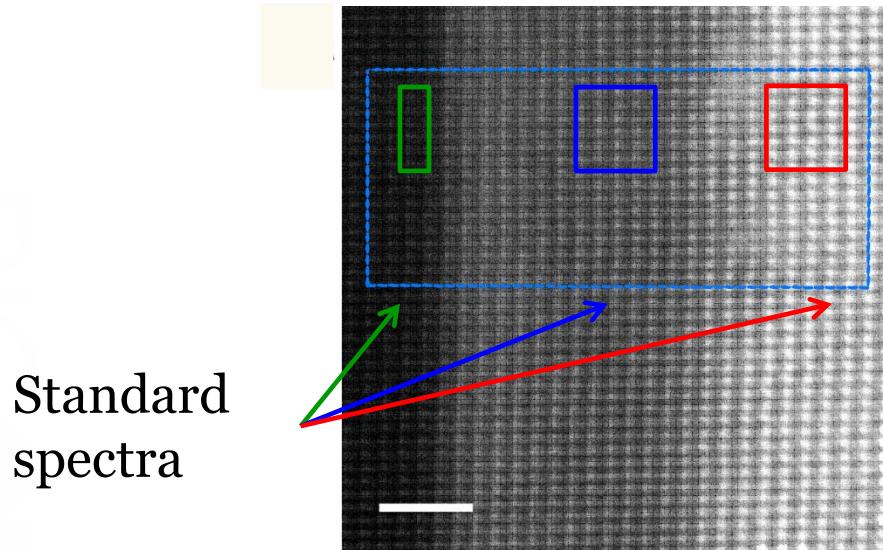
Overall, we can have (maximum) P loading maps of M^2 size. However, not all of them will have useful information



Multiple Linear Regression

Linear mixing $S(\mathbf{x}, \mathbf{R}) = \sum_i a_i(\mathbf{x}) w_i(\mathbf{R}) + N$ but $w_i(\mathbf{R})$ are known

STEM of STO/LSMO/BFO interface Low-loss EELS spectra of three components



A.Y. BORISEVICH ET AL,
Suppression of Octahedral Tilts and Associated Changes in Electronic Properties at Epitaxial Oxide Heterostructure Interfaces, Phys. Rev. Lett. **105**, 087204 (2010).

“Chemistry”:
35 to 125 eV



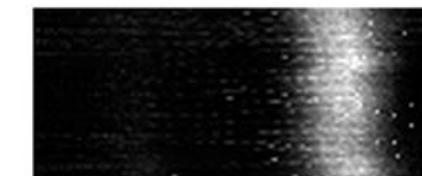
“Plasmons”
5 to 35 eV



Fit coefficient map

residuals map

χ^2 map



EELS elemental mapping with unconventional methods I. Theoretical basis: image analysis with multivariate statistics and entropy concepts

Pierre Trebbia *

Laboratoire de Physique des Solides, Bâtiment 510, F-91405 Orsay Cedex, France

and

Noël Bonnet

Unité INSERM 314 et Université de Reims, 21 rue Clément Ader, F-51100 Reims, France

Received 7 June 1990

Electron energy loss filtered images recorded within a transmission analytical electron microscope are now widely used for the mapping of the elemental distribution of a given atomic species in a specimen prepared as a thin film. Such an image processing may produce both valuable results and artifacts if a careful inspection of all the hypotheses needed by the calculation is not carried out. This paper presents some general statistical methods for a contrast information analysis of a noisy image data set. After a brief introduction of different concepts such as contrast, variance, information and entropy, two unconventional approaches for image analysis are explained: the relative entropy computed with respect to a pure random and signal-free image and the factorial analysis of correspondence (a branch of multivariate statistics). In the companion article (part II), these concepts are applied to real experiments and the results compared with those obtained with a conventional method. Although electron energy loss spectroscopy is the only technique considered here, these methods for image analysis can be applied to a wide variety of noisy data sets (spectra, images, ...) recorded from various sources (electrons, photons, ...).

Why historical papers matter:

- 165 1. Often contain elementary introductions
2. Deep insights into principles
3. Surprisingly prescient predictions
4. Comparison with the present: see the big picture

“Those who cannot remember the past are condemned to repeat it.”

George Santayana,
The Life of Reason, 1905



Available online at www.sciencedirect.com



Ultramicroscopy 106 (2006) 1024–1032

ultramicroscopy

www.elsevier.com/locate/ultramic

Mapping chemical and bonding information using multivariate analysis of electron energy-loss spectrum images

M. Bosman^{a,*}, M. Watanabe^b, D.T.L. Alexander^c, V.J. Keast^a

^aAustralian Key Centre for Microscopy and Microanalysis, University of Sydney, Sydney, NSW 2006, Australia

^bDepartment of Materials Science and Engineering, Lehigh University, Bethlehem, PA 18015, USA

^cDepartment of Materials Science and Metallurgy, University of Cambridge, Pembroke St. CB2 3QZ, UK

Received 23 June 2005; received in revised form 26 October 2005; accepted 18 April 2006

Abstract

Electron energy-loss spectroscopy (EELS) in the transmission electron microscope (TEM) is used to obtain high-resolution information on the composition and the type of chemical bonding of materials. Spectrum imaging, where a full EEL spectrum is acquired and stored at each pixel in the image, gives an exact correlation of spatial and spectral features. However, determining and extracting the important spectral components from the large amount of information contained in a spectrum image (SI) can be difficult. This paper demonstrates that principal component analysis of EEL SIs can be used to extract chemically relevant components. With weighted or two-way scaled principal component analysis, both compositional and bonding information can be extracted. Mapping of the chemical variations in a partially reduced titanium dioxide sample and the orientation-dependent bonding in boron nitride and carbon nanotubes are given as examples.

© 2006 Elsevier B.V. All rights reserved.

Why did the PCA on EELS started to grow in 2005 – 2010?

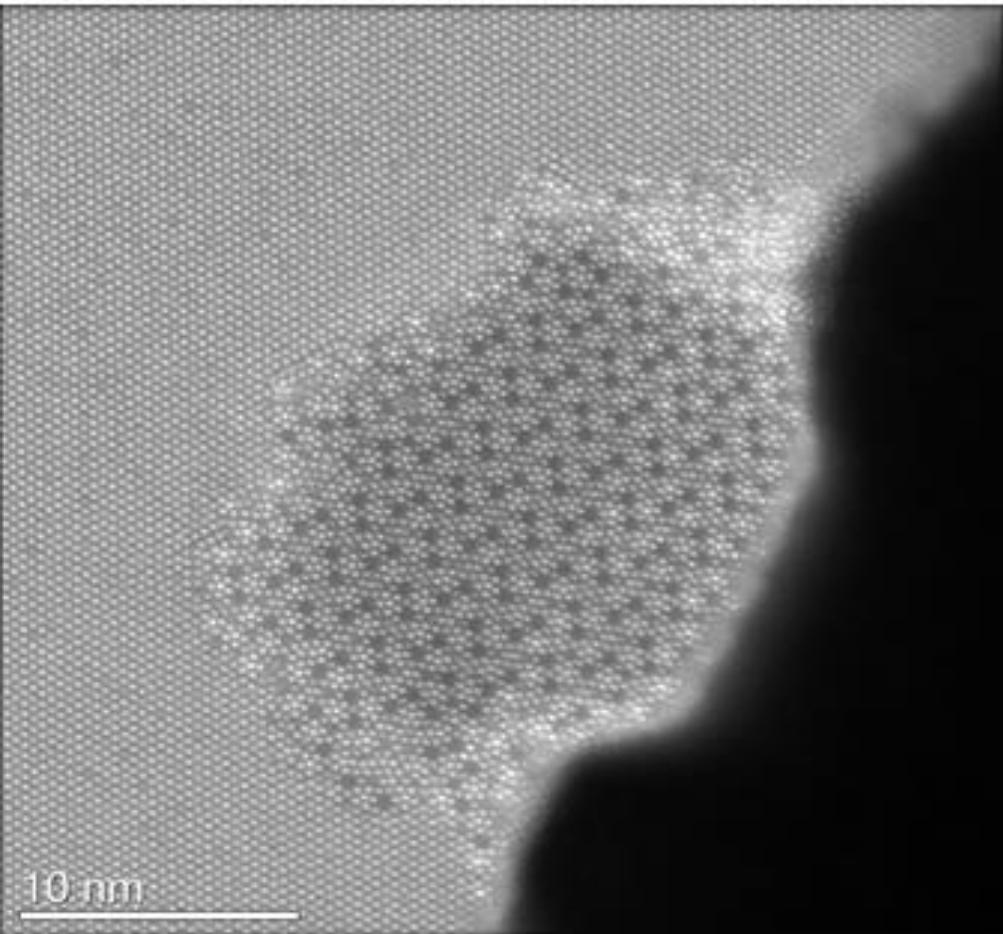
Dimensionality Reduction Methods

- Multiple Linear Regression
 - Present data as a sum of known components
- PCA (Principal Component Analysis):
 - Find projection that maximize the variance
- ICA (Independent Component Analysis):
 - Very similar to PCA except that it assumes non-Gaussian features
- Multidimensional Scaling:
 - Find projection that best preserves inter-point distances
- LDA(Linear Discriminant Analysis):
 - Maximizing the component axes for class-separation
- Bayesian Linear Unmixing
 - Linear unmixing, non-negative, sum to one
- ... constrained linear unmixing methods

Colab

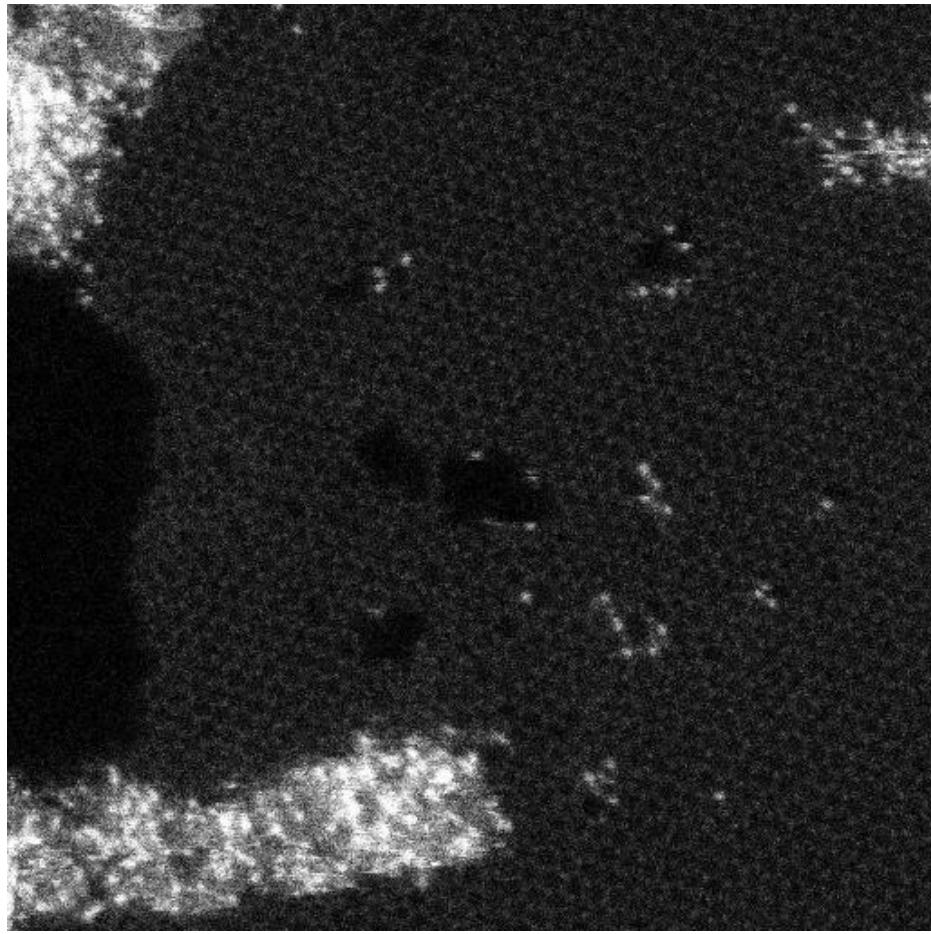
Chemically disordered systems

Mo-V-Ta complex oxide



Q. He et al, ACS Nano 9, 3470-3478

Si in graphene



Data collected by O. Dyck (ORNL)

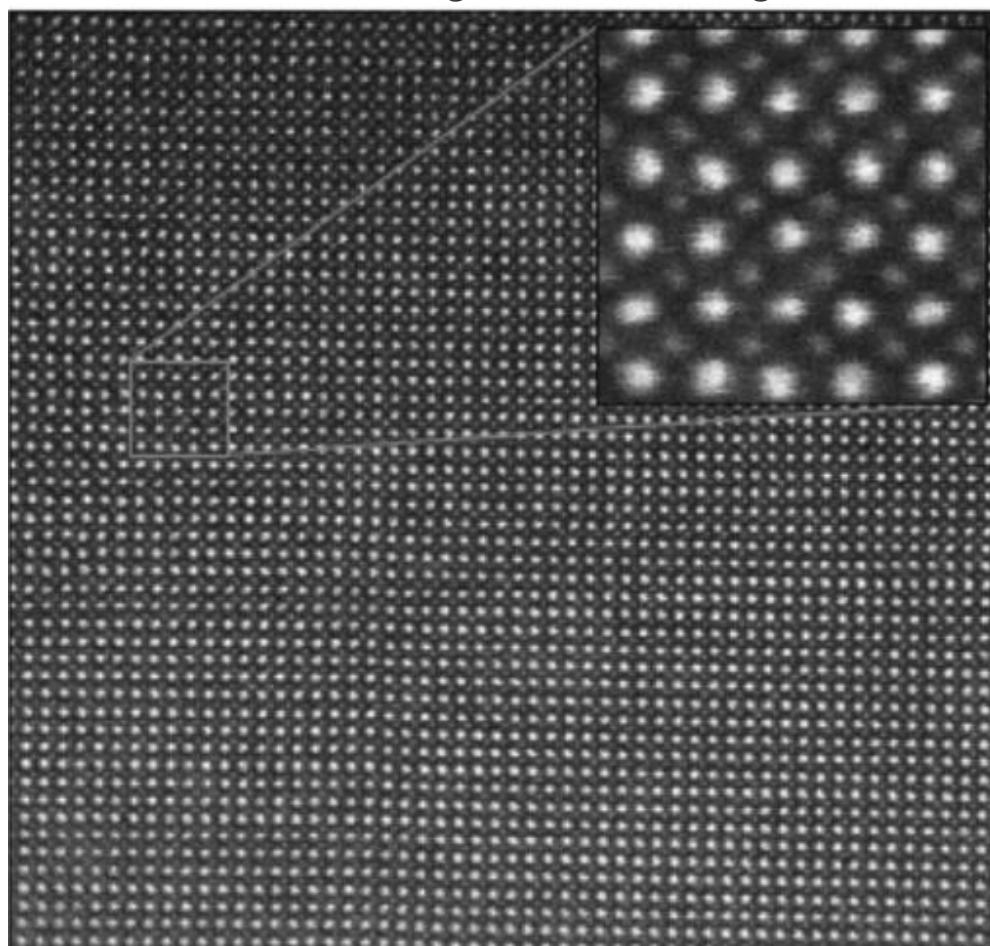
- What is the nature of the building blocks and relevant atomic configurations?
- Can we define single-phase regions and phase boundaries?

But what about subtle distortions?

Electronic structure in RuCl₃



BiFeO₃ on SrRuO₃



- Can we identify ferroelectric and ferroic variants and associated topological defects?
- What is the nature of the phases?

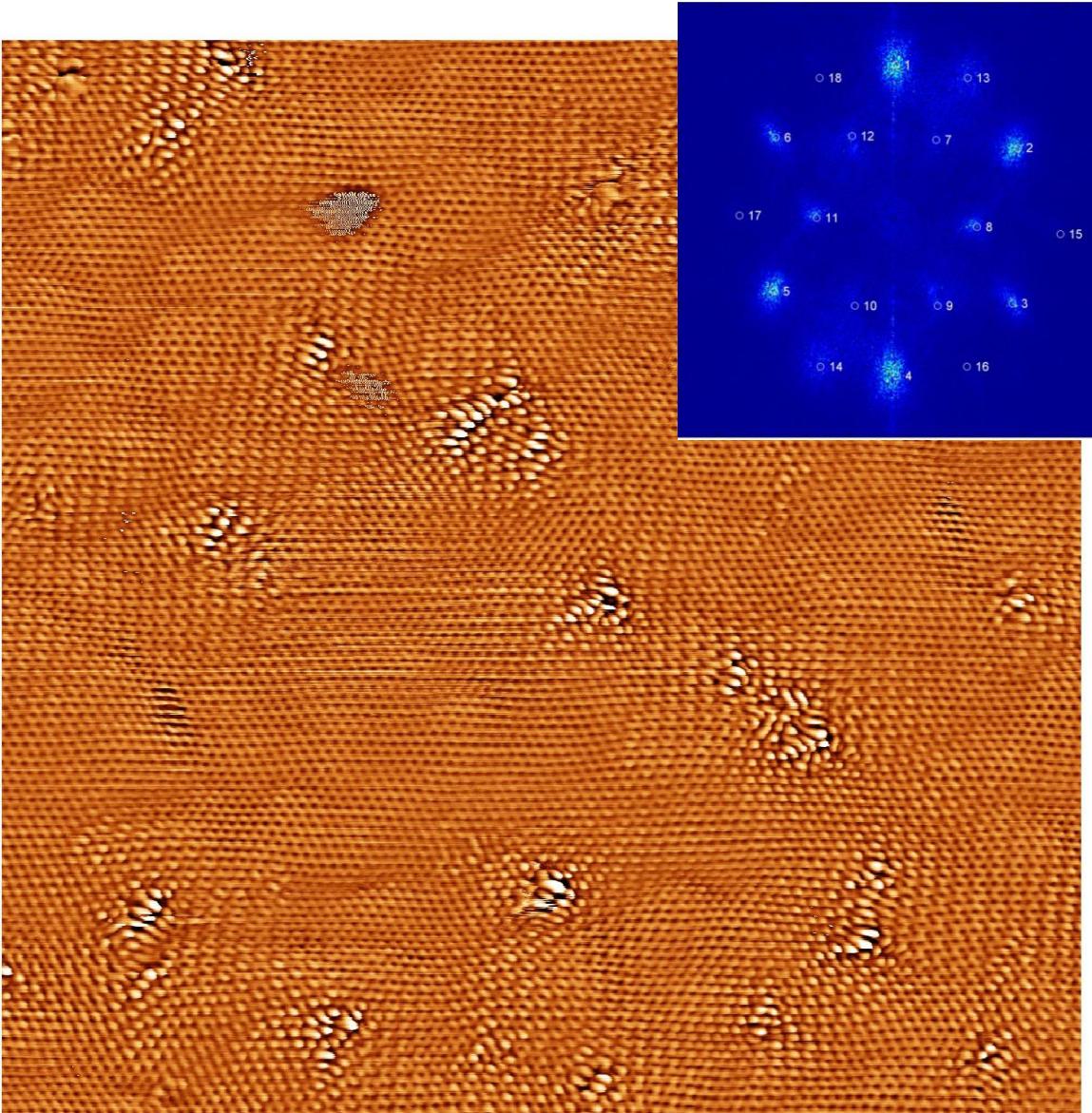
Can global FFT help?

- Global FFT – everything is averaged:
 - Drift
 - Extended defects
 - Multiple grains

We are averaging out all interesting phenomena except for small spatially uniform structural distortions

- Solution – sliding window approaches:
 - Fit FFT peaks: amplitudes, positions
 - Multivariate analysis

Note: window can be also tied to a specific feature, such a selected atom. Then we explore atomic neighborhood



Sliding FFT:

- We always have a problem of window size:
 - too large – loose spatial resolution,
 - too small – FFT behaves poorly due to edge effects
- Interpretation of FFT data is complicated (too much data if fit each peak, unclear meaning of the unmixing components)
- Natural descriptor for atomically resolved images – atomic coordinates!

General linear unmixing

$$S(\mathbf{x}, \mathbf{R}) = \sum_i a_i(\mathbf{x}) w_i(\mathbf{R}) + N$$

We start with:

- \mathbf{x} is the spatial variable, $\mathbf{x} = (x, y)$
- \mathbf{R} is the (vector) parameter variable

We aim to get:

- $a_i(\mathbf{x})$ are loading maps
- $w_i(\mathbf{R})$ are endmembers/eigenvectors
- N is noise

The M pixel 2D image is transformed to M/N pixel image of more complex structure.

Our loading map is 2D image, and
endmembers/eigenvectors are 2D images

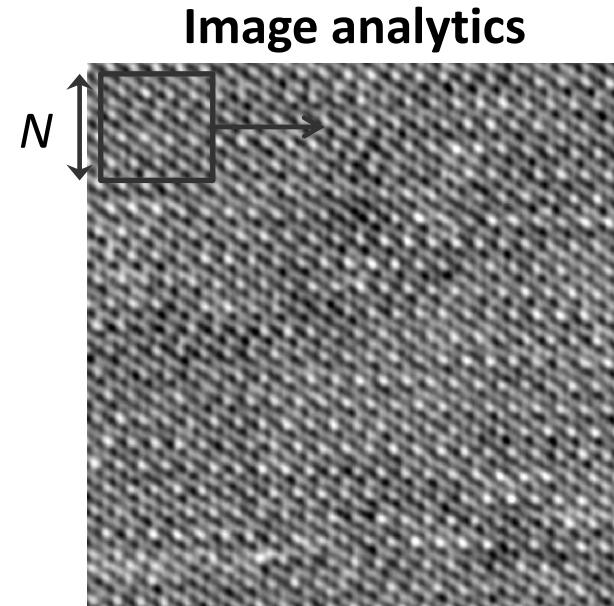


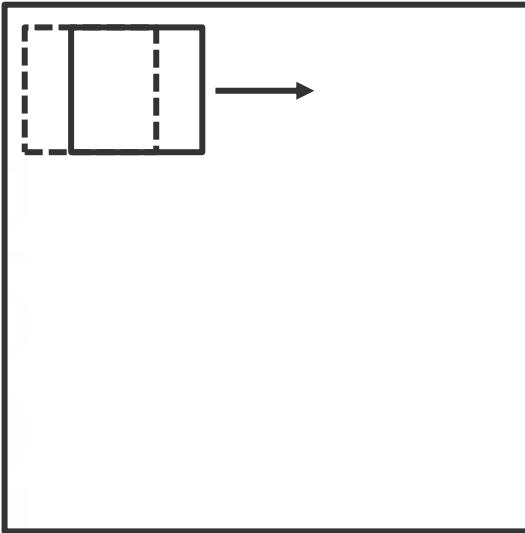
Figure by M. Ziatdinov

Sliding image transforms:

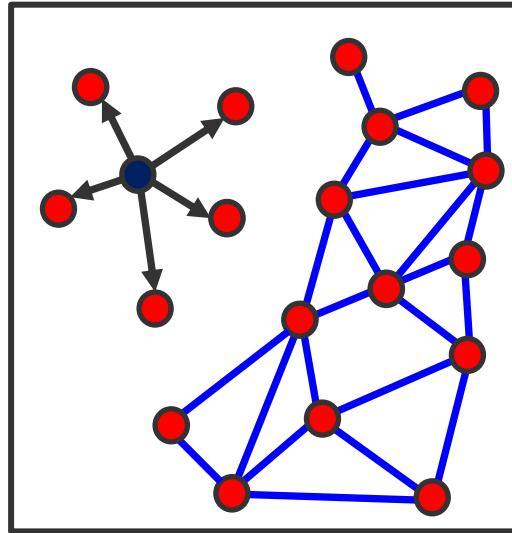
- Fast Fourier Transforms
- Correlation functions
- Intensity histograms
- Structural descriptors

Constructing the descriptors

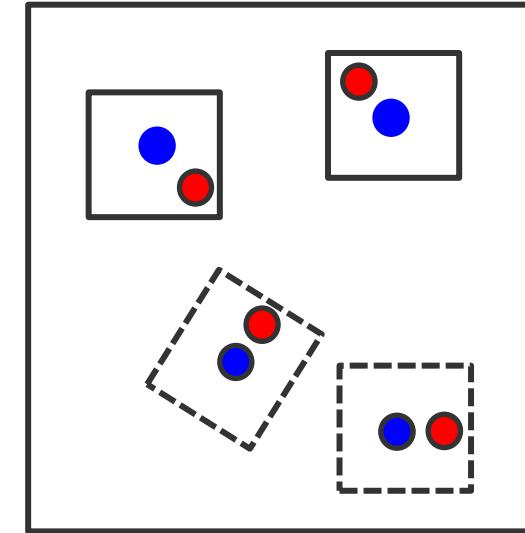
**Continuous
translational
symmetry**



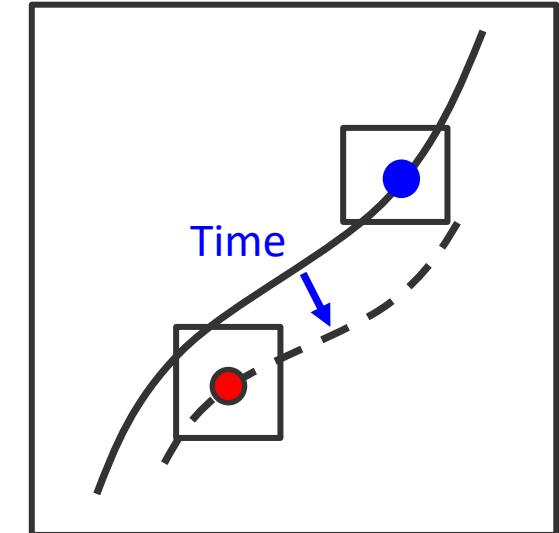
**Atom based
descriptions**



**Localized
sub-images**



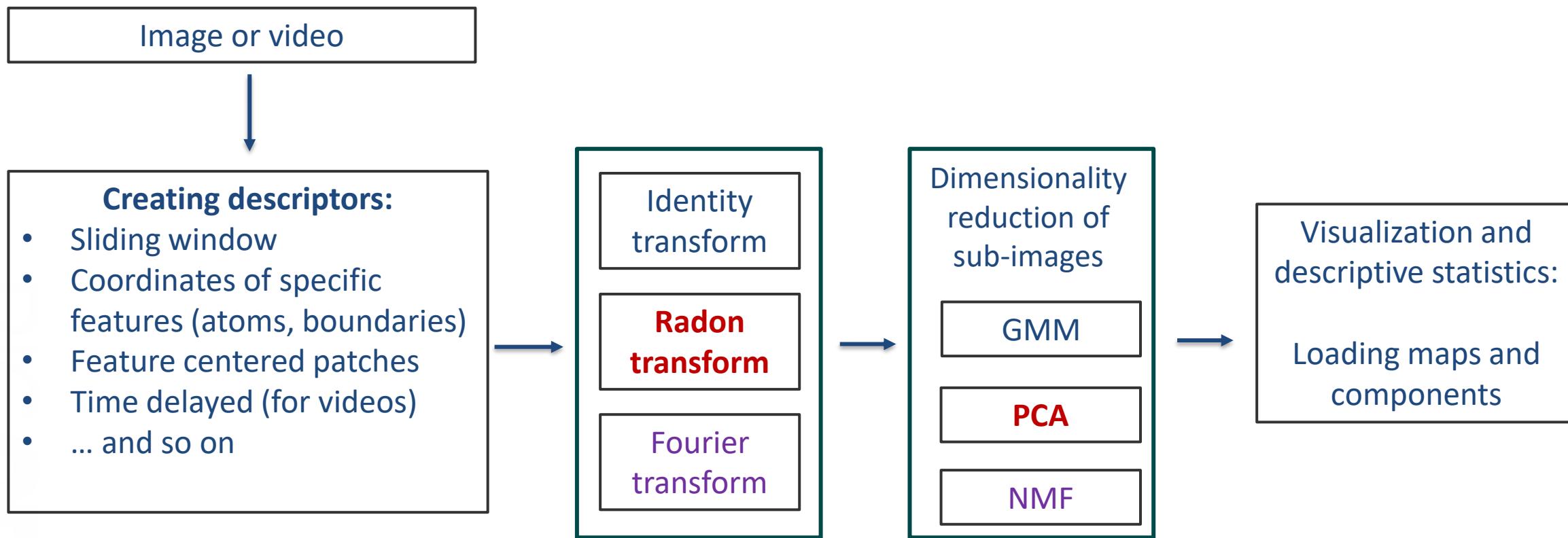
**Time-delayed
descriptors**



The choice of the descriptor:

- Defines physical inferential biases and allows to introduce prior knowledge
- Determines the physical meaning of the analysis
- Establishes the analysis pipeline

Example of analysis pipeline



Pipelines are defined to

- Make analysis traceable, repeatable, explainable, and transferable
- Allow for hyperparameter tuning and optimization
- Efficiently use the memory

How general should you be: depends on applications

Principal Component Analysis

$$S(\mathbf{x}, \mathbf{R}) = \sum_i a_i(\mathbf{x}) w_i(\mathbf{R})$$

In PCA, the eigenvectors $w_i(\mathbf{R})$ are orthonormal and are arranged such that corresponding eigenvalues are placed in descending order by variance

- Reveals internal structure of the data that best explains variance in the data set
- Since data often moves in clusters, PCA reveals those variables that drive the variance
- PCA transforms the data such that the greatest variance by any projection lies on the first coordinate

Sliding PCA-FFT

Can we use PCA of FFT transform in sliding windows to find periodicity?

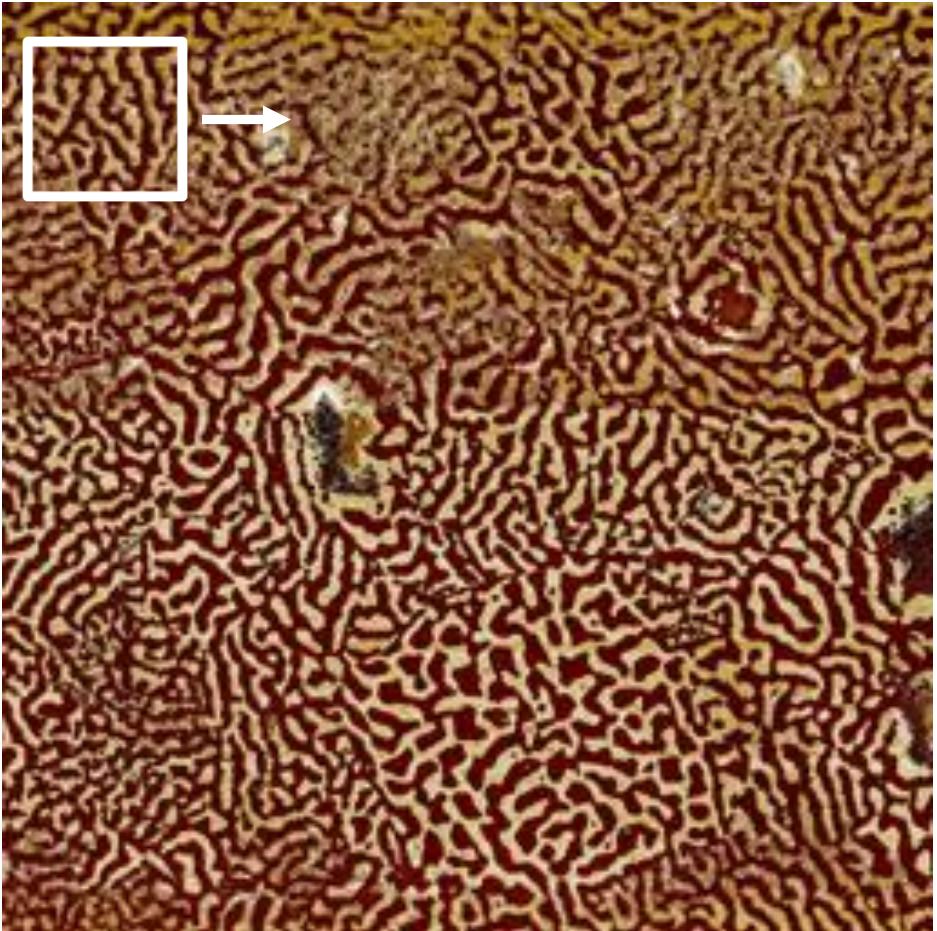
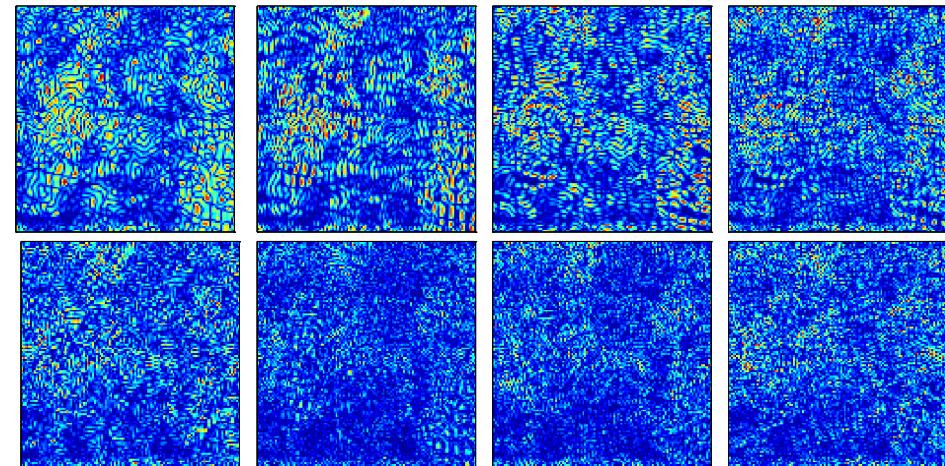
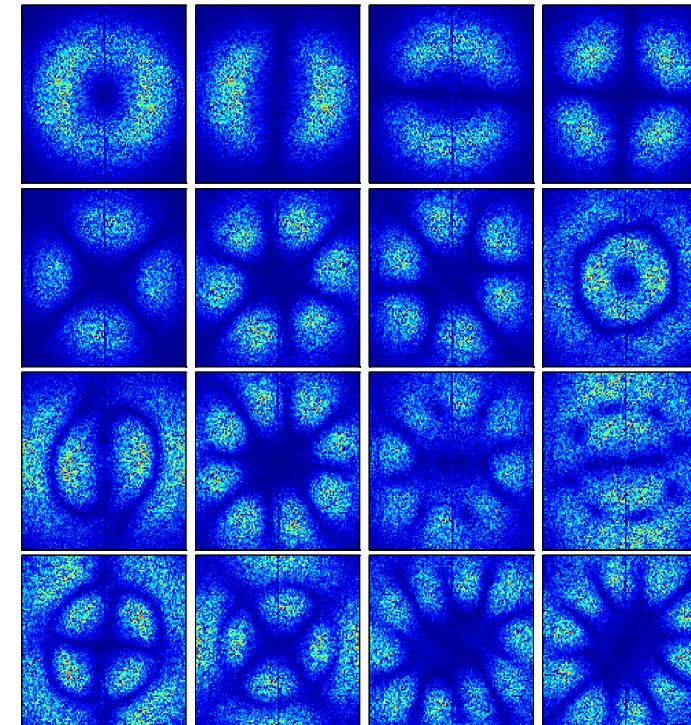


Figure by S. Jesse, data D. Gobellic

First 8 maps



First 16 eigenvectors



Spectral Unmixing: N-FINDR

Spectra for a given pixel is assumed to be a linear combination of the end-member spectra (+ Gaussian noise). The mixing proportions sum to 1

Physics constraint

$$p_{ij} = \sum_k e_{ik} c_{kj} + \varepsilon \quad \sum_k c_{kj} = 1$$

- Let E be the matrix of end-members (here, 3).

$$E = \begin{bmatrix} 1 \\ \vec{e}_1 & \vec{e}_2 & \vec{e}_3 \end{bmatrix} \quad V \left(\frac{1}{(l-1)!} \right) |\det(E)|$$

- Iteratively select endmembers, accepting the new selection if the volume increases

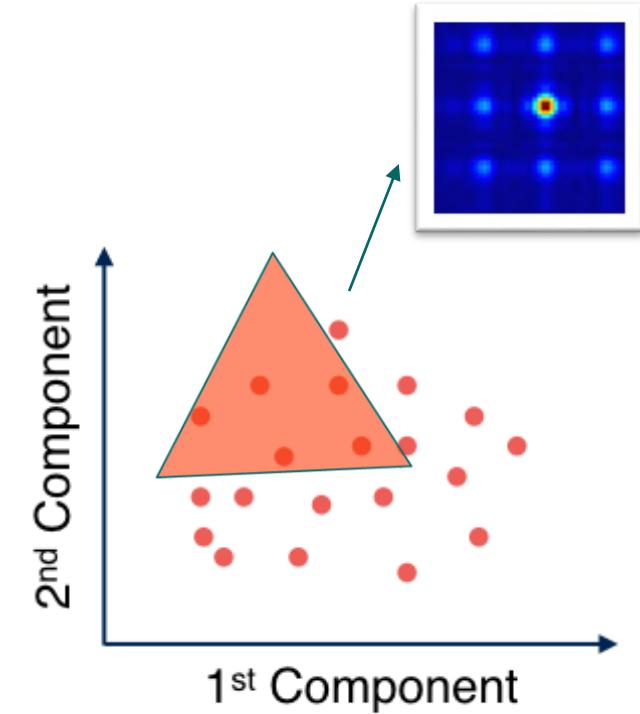


Figure by R. Vasudevan

Ideal test case

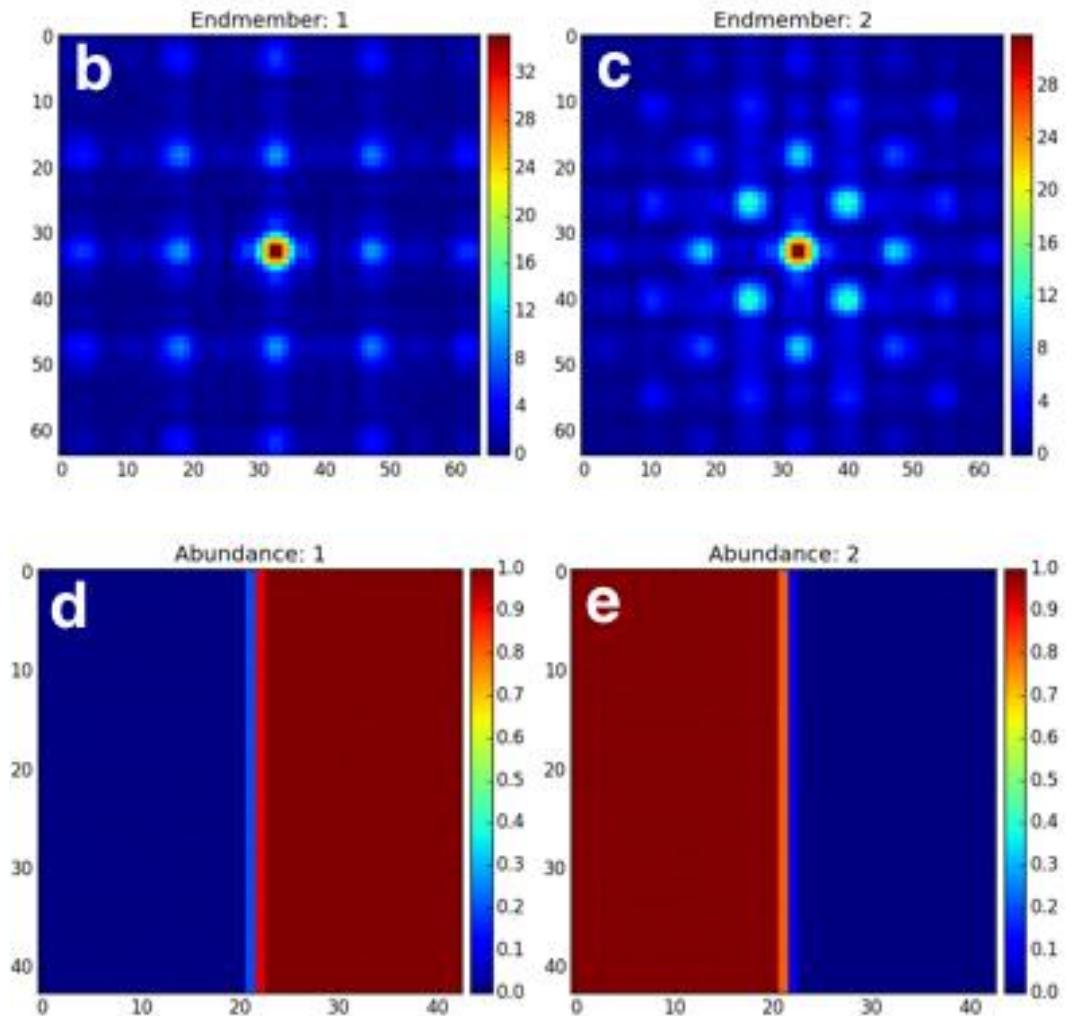
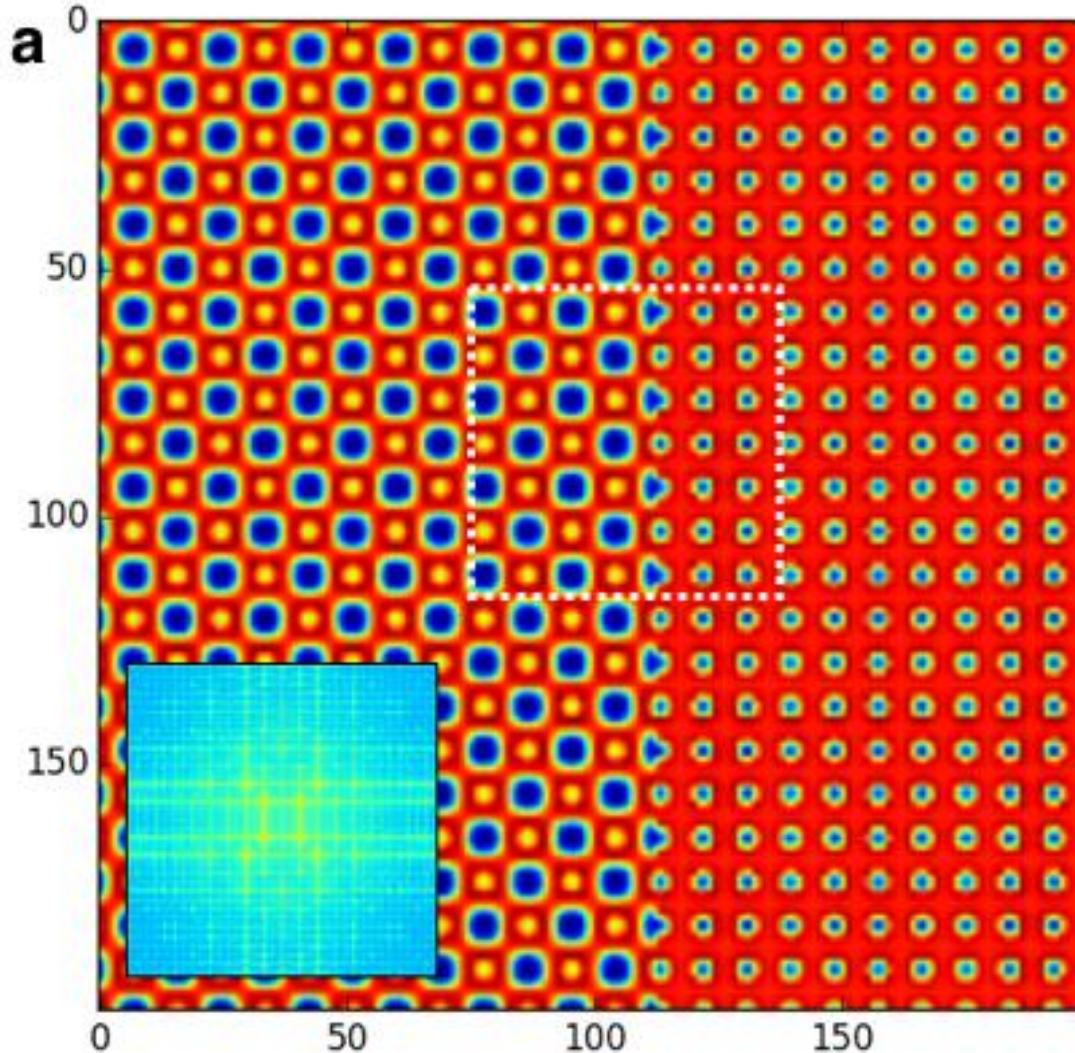
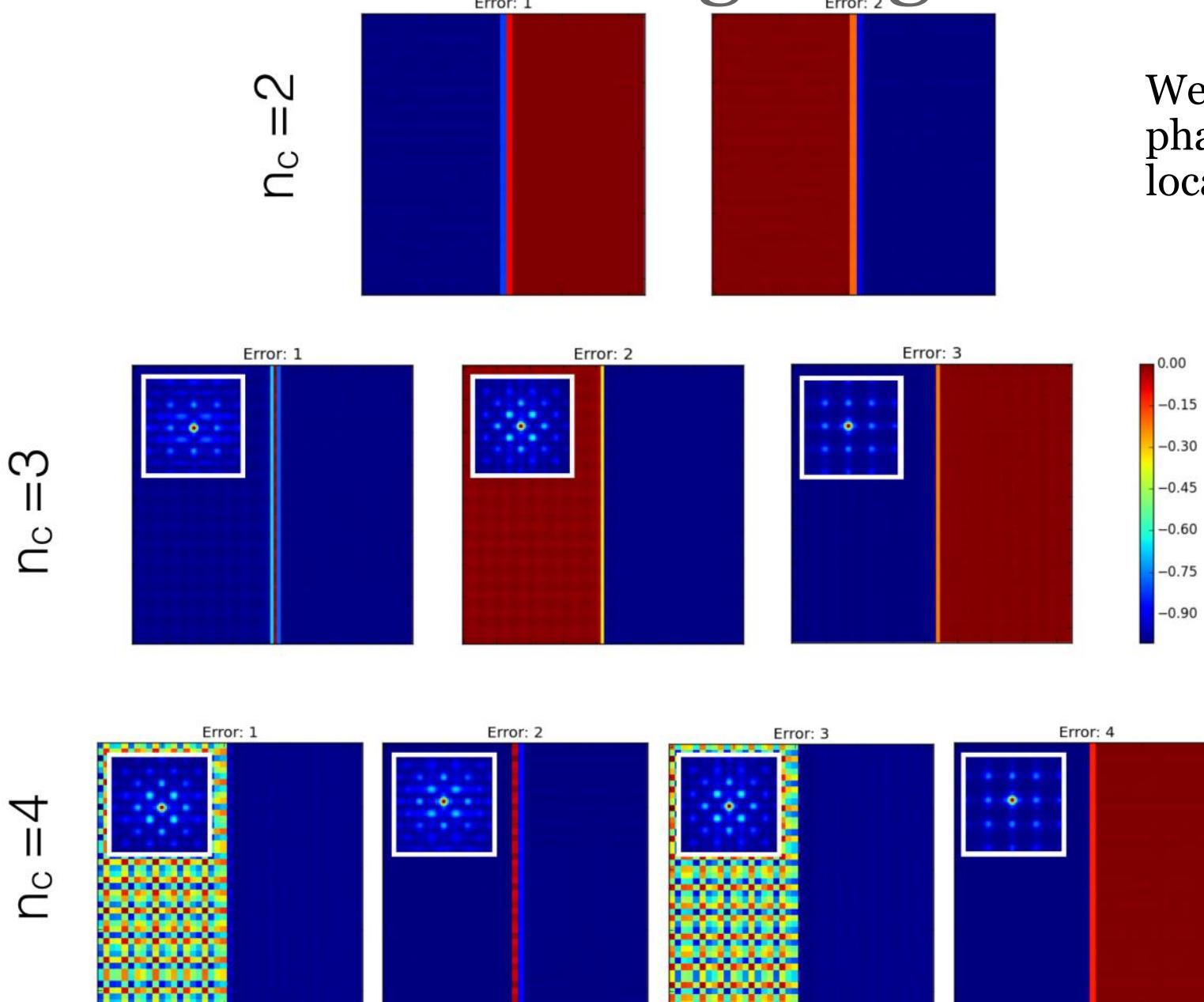


Figure by R. Vasudevan

Main idea:

- FFT amplitudes are non-negative;
- FFT removes translation

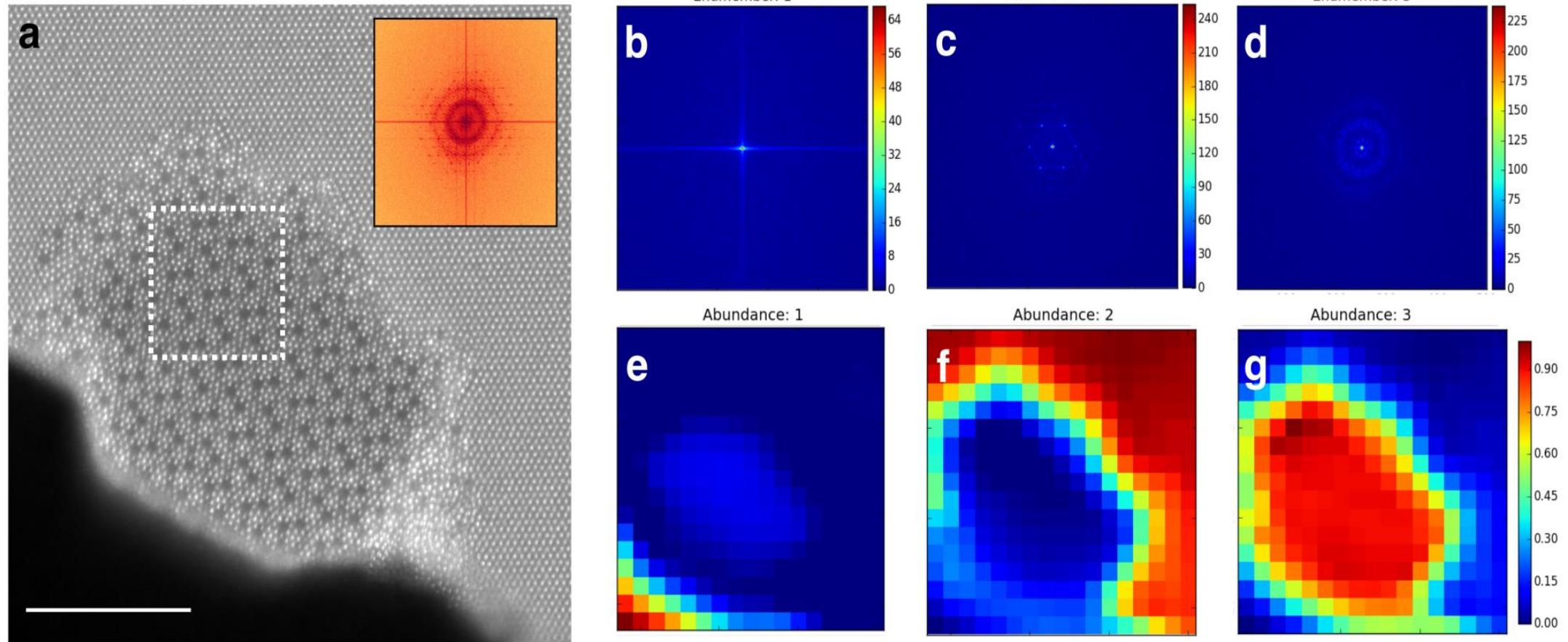
N-FINDR for image segmentation



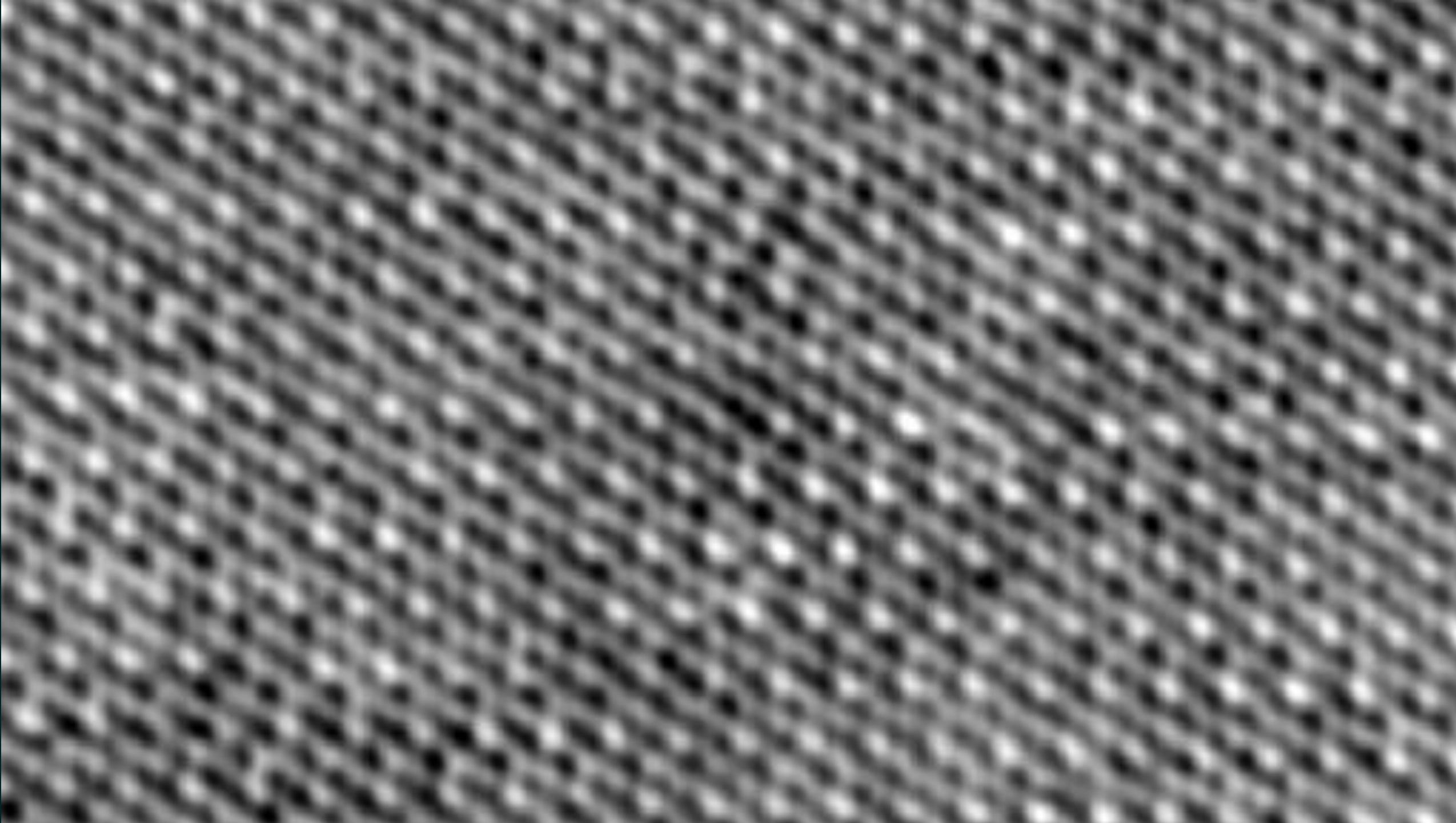
We can determine number of phases based on spatial localization and geometry

Figure by R. Vasudevan

N-FINDR for chemically separated images

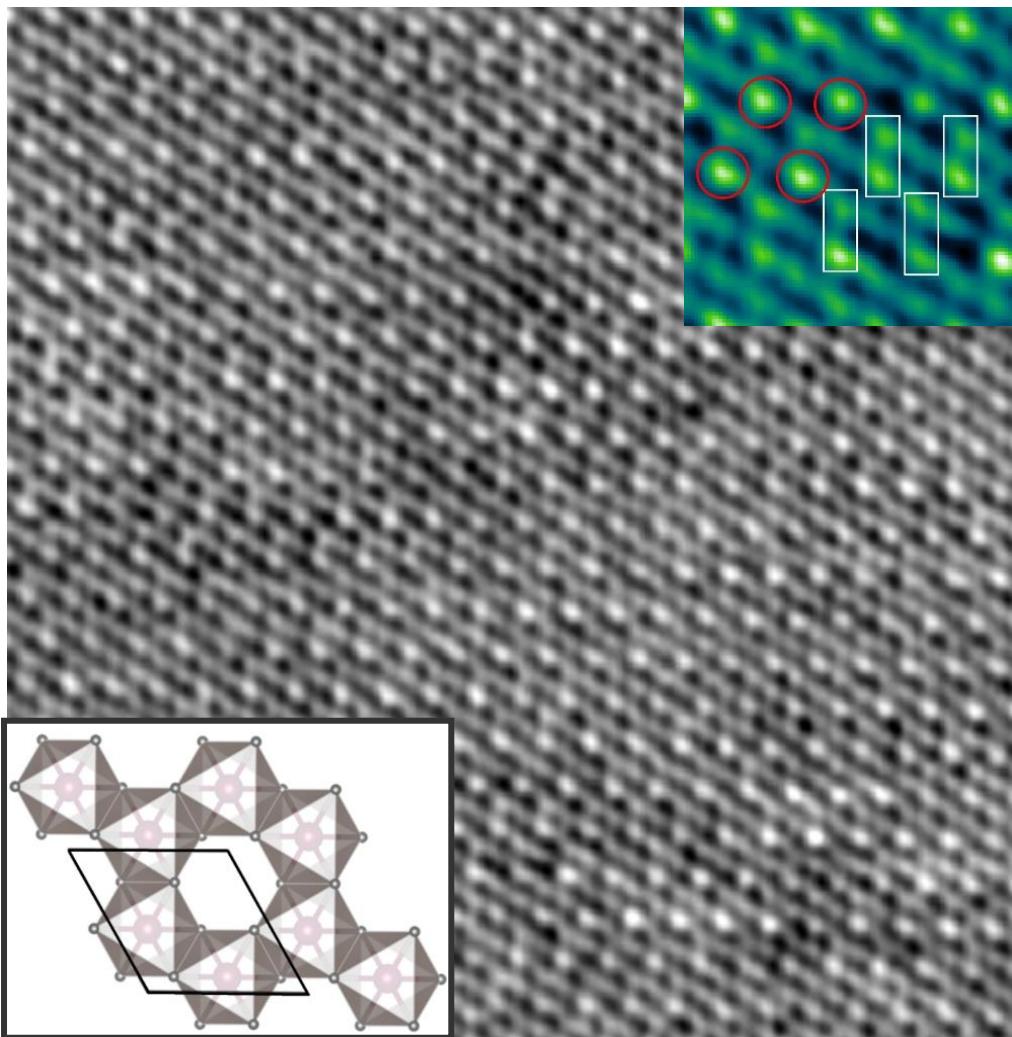


Q. He, J. Woo, A. Belianinov, V.V. Gulians, A.Y. Borisevich, *Better catalysts through microscopy: mesoscale M₁/M₂ intergrowth in Molybdenum–Vanadium based complex oxide catalysts for propane ammoxidation*, ACS Nano 9, 3470-3478

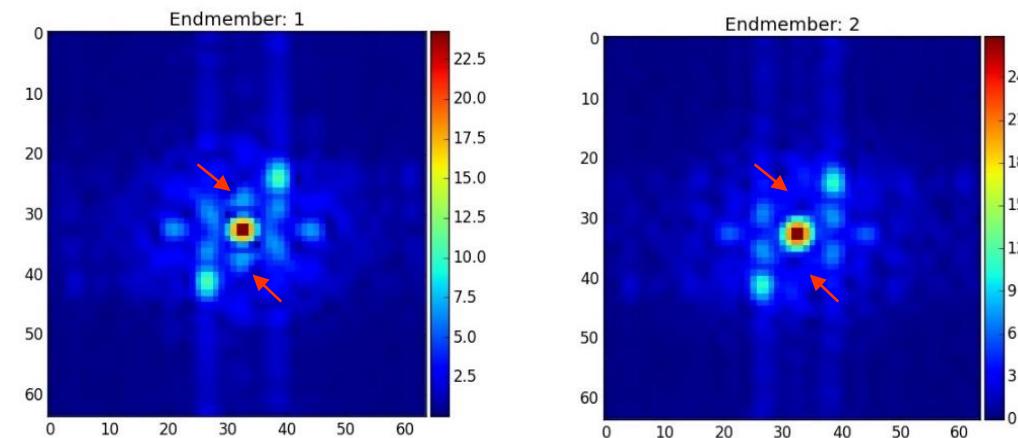


NFINDR for coexisting order parameters

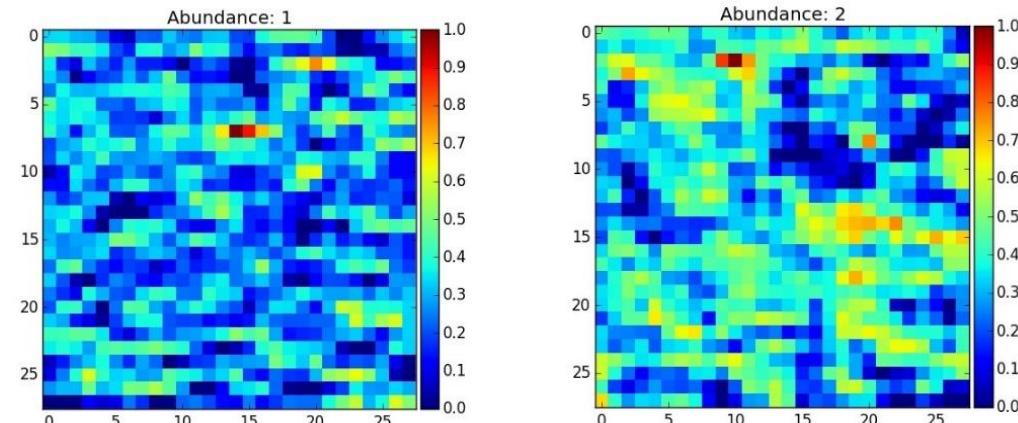
Input experimental image



FFT endmembers



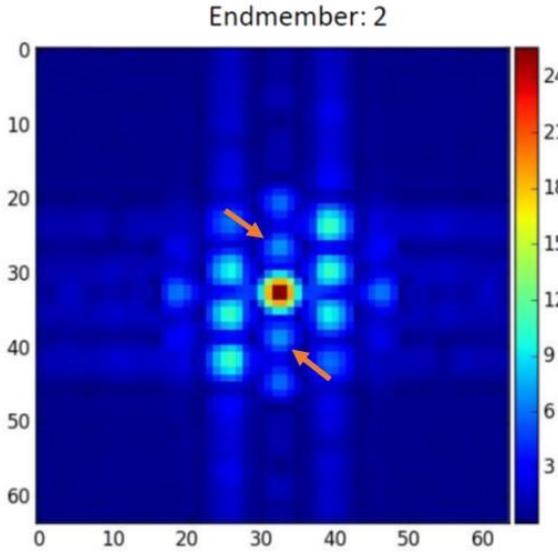
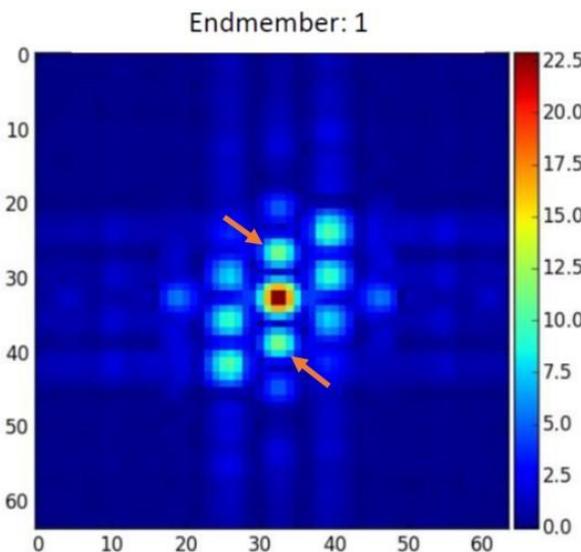
Real-space abundance maps



In a good agreement with test case, 2 spots in the “inner hexagon” are strongly suppressed in the 2nd component reflecting a fine structure of charge ordered pattern

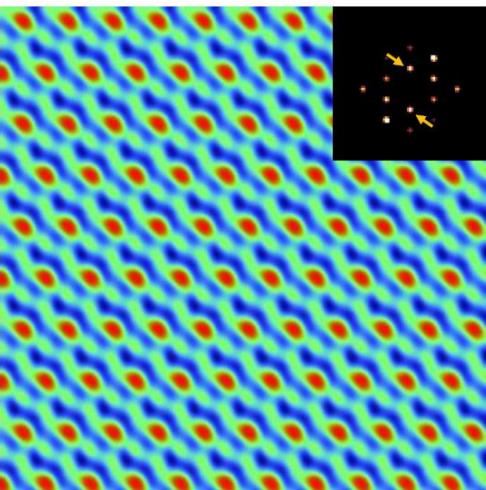
NFIND-R for coexisting order parameters

FFT endmembers

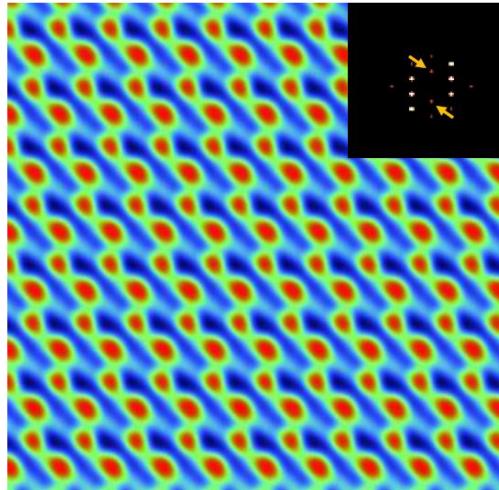


In the 2nd component, 2 spots in the “inner hexagon” are strongly suppressed reflecting a fine structure of charge ordered pattern

Real-space images of corresponding phases



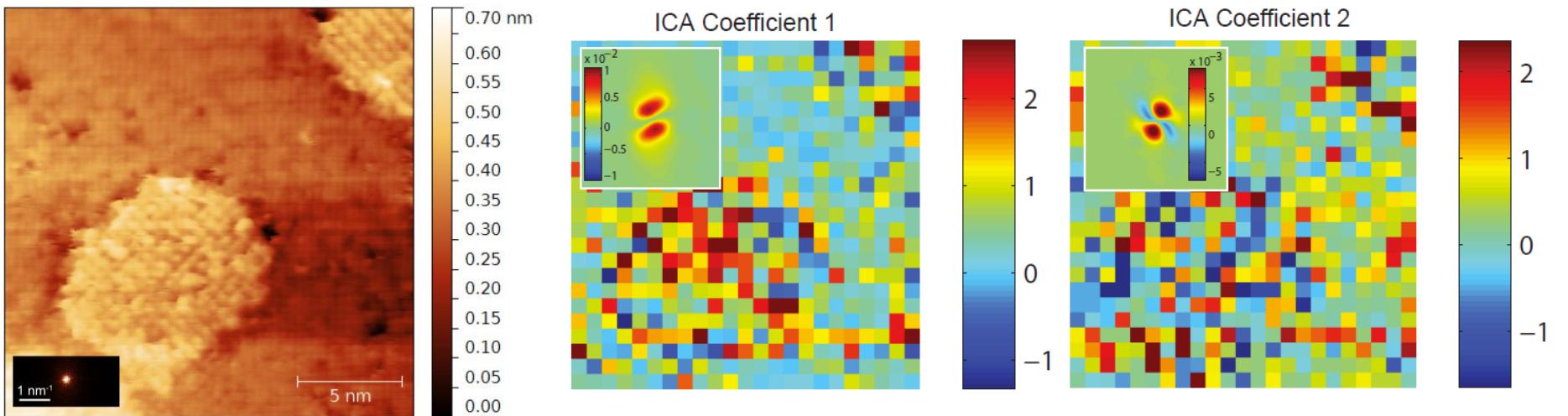
Hexagonal superlattice



Dimer superlattice

M. ZIATDINOV, A. BANERJEE, A. MAKSOV, T. BERLIJN, W. ZHOU, H.B. CAO, J.Q. YAN, C.A. BRIDGES, D.G. MANDRUS, S.E. NAGLER, A.P. BADDORF, and S.V. KALININ, *Atomic-scale observation of structural and electronic orders in the layered compound α -RuCl₃*, Nature Comm. 7, 13774 (2016).

ICA on Sliding FFT



- Appears to separate into pairs of components
- Unsuitable to the physics of the problem

Figure by R. Vasudevan

NFIND-R vs. ICA

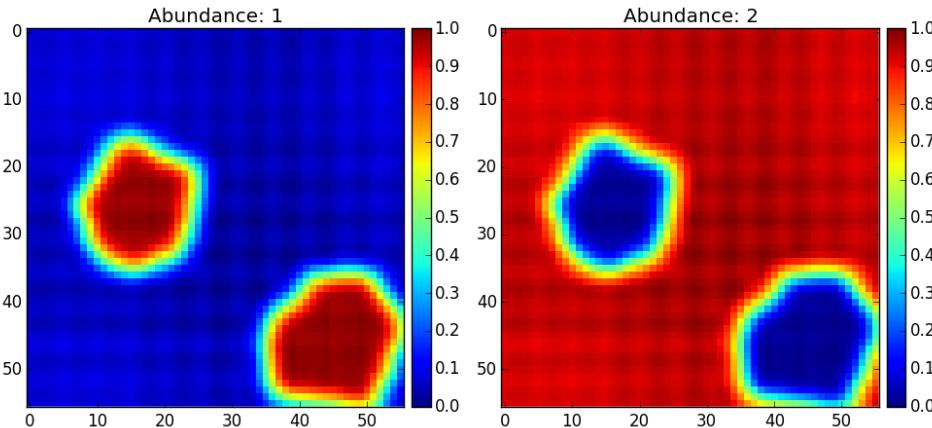
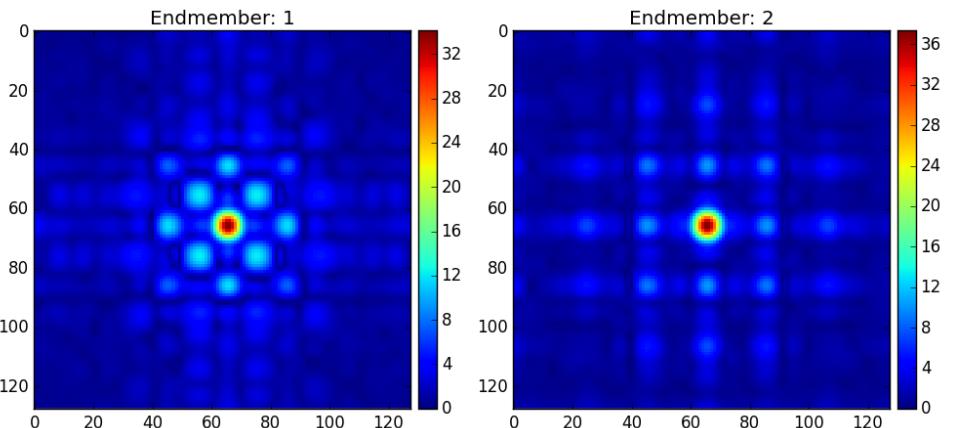
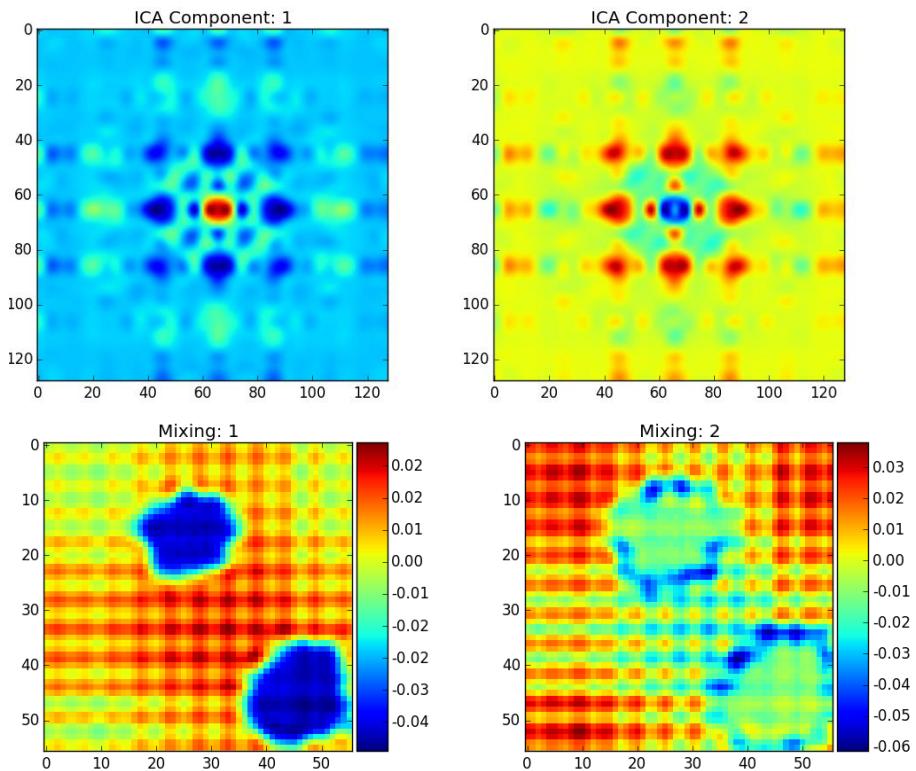
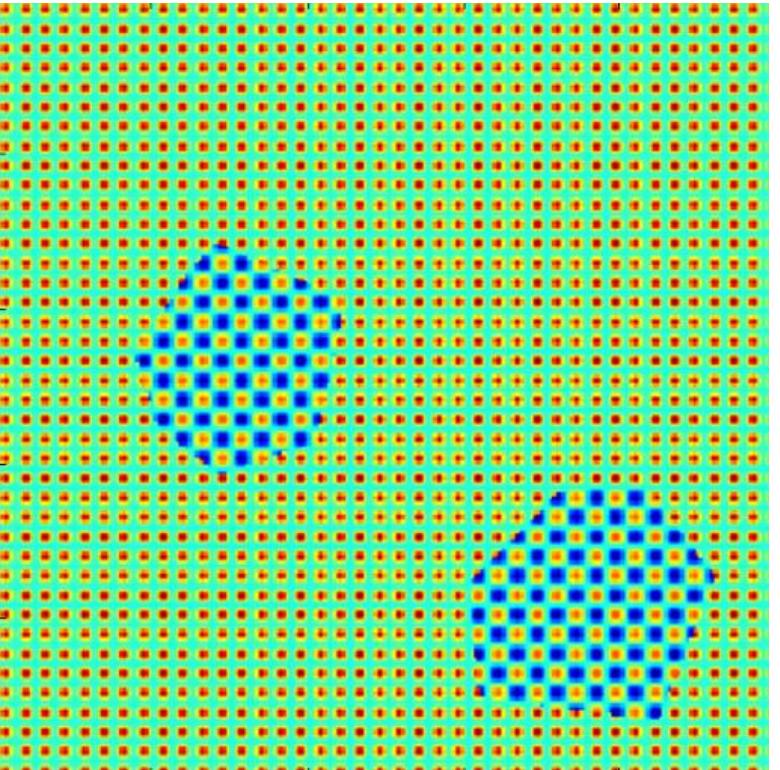


Figure by
R. Vasudevan

Sliding FFT:

- We always have a problem of window size:
 - too large – loose spatial resolution,
 - too small – FFT behaves poorly due to edge effects
- Interpretation of FFT data is complicated (too much data if fit each peak, unclear meaning of the unmixing components)
- Natural descriptor for atomically resolved images – atomic coordinates!

Local crystallography

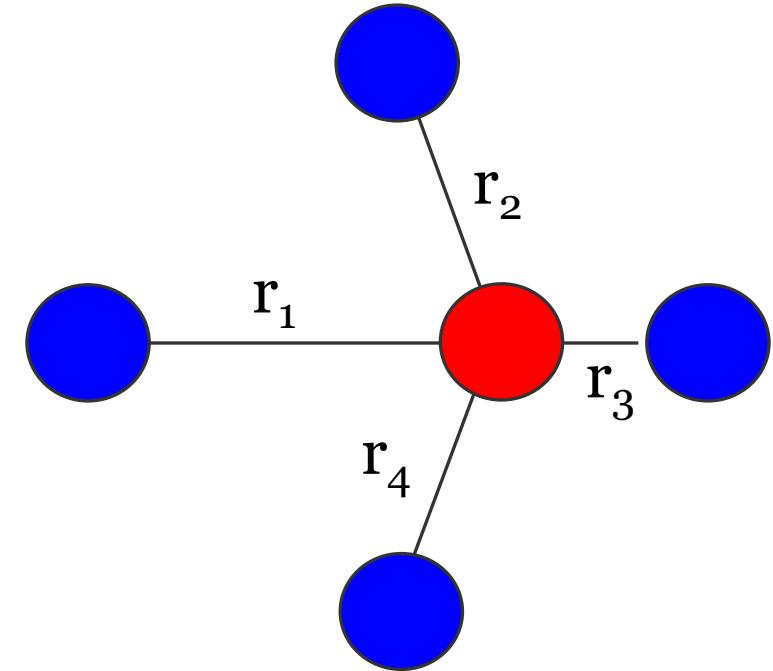
For each atom, define nearest neighbors and generate array of the corresponding radius-vectors of the form

$$NA_{ij} = (rx_1, ry_1, rx_2, ry_2, rx_3, ry_3, rx_4, ry_4)_{ij}$$

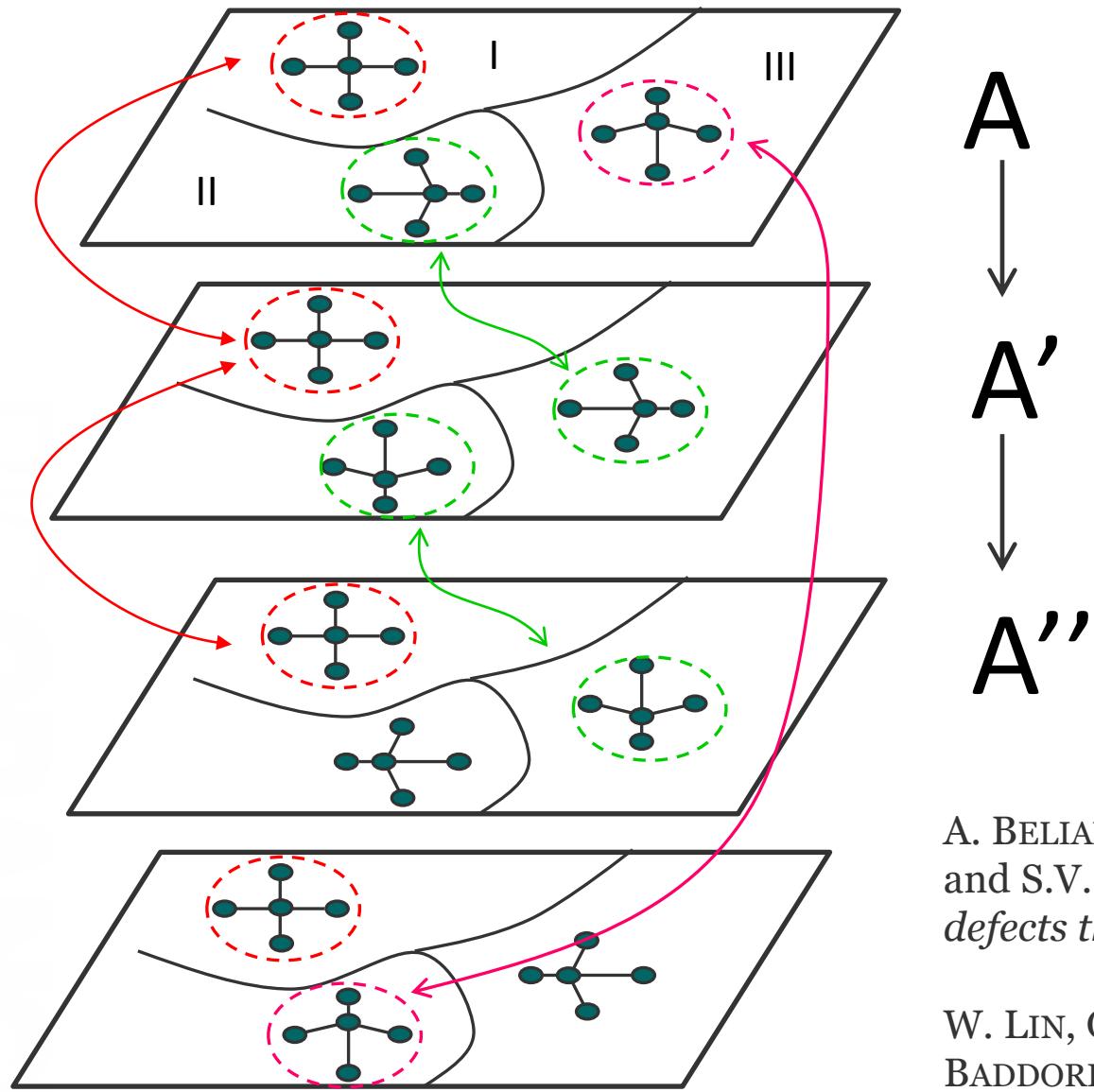
Indexes 1,2,3,4 are chosen in the same sense for all atoms
(generalization for different lattice and/or next coordination sphere obvious)

Then, phase/ferroic variant identification problem can be reduced to finding equivalent (in statistical sense) groups of nearest neighbors

We can also use group theory to make hypotheses, e.g.
add translation symmetry operations, i.e. $i \rightarrow i+1$ and $j \rightarrow j+1$ for lattice doubling)



Local crystallography



Same cluster in all replicas:
Non-ferroic phase

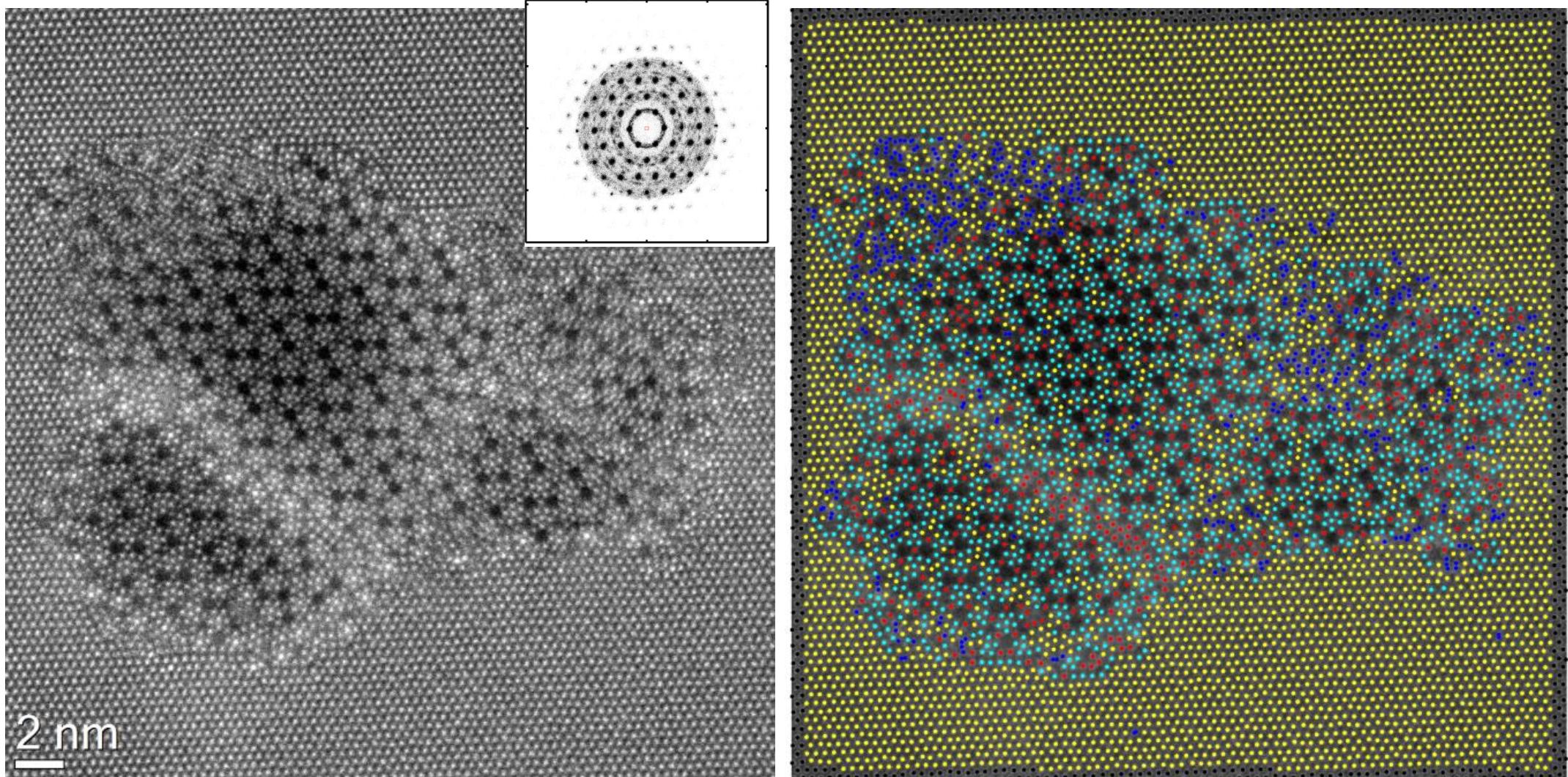
Form cluster with different regions
in different replicas:
Ferroic phase

Only some of the correspondences
are shown (but these are obvious)

A. BELIANINOV, Q. HE, M. KRAVCHENKO, S. JESSE, A. BORISEVICH,
and S.V. KALININ, *Identification of phases, symmetries, and
defects through local crystallography*, Nat. Comm. **6**, 7801 (2015).

W. LIN, Q. LI, A. BELIANINOV, B.C. SALES, A. SEFAT, Z. GAI, A.P.
BADDORF, M. PAN, S. JESSE, and S.V. KALININ, *Local
crystallography analysis for atomically resolved scanning
tunneling microscopy images*, Nanotechnology **24**, 415707 (2013).

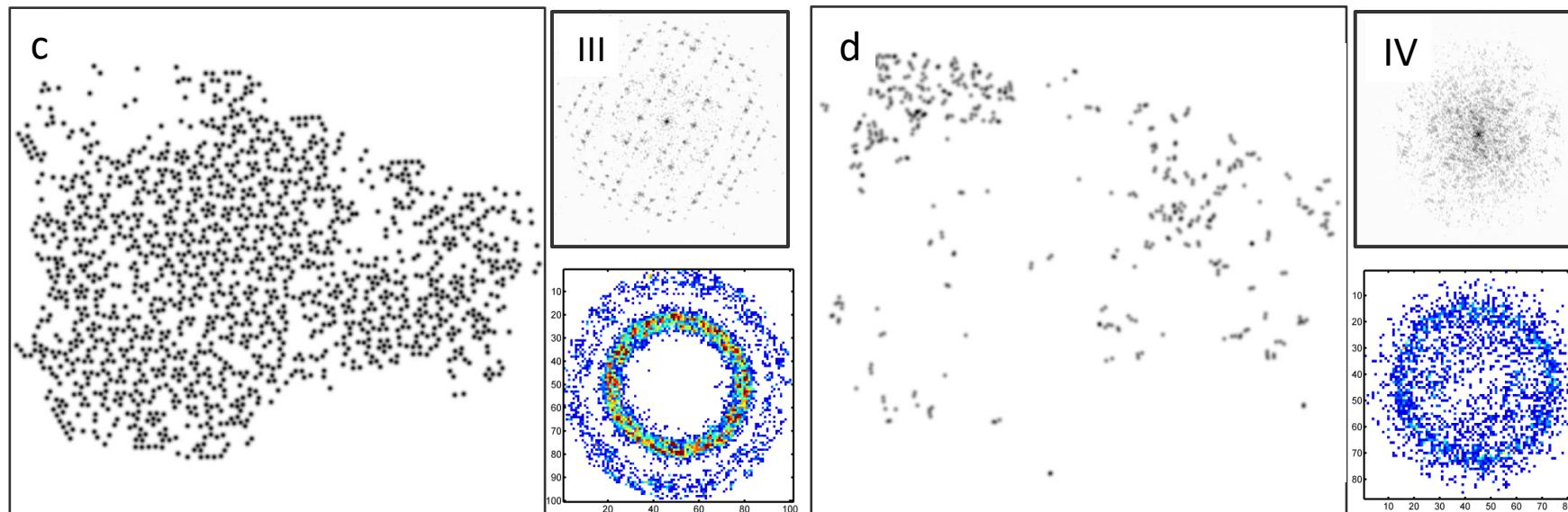
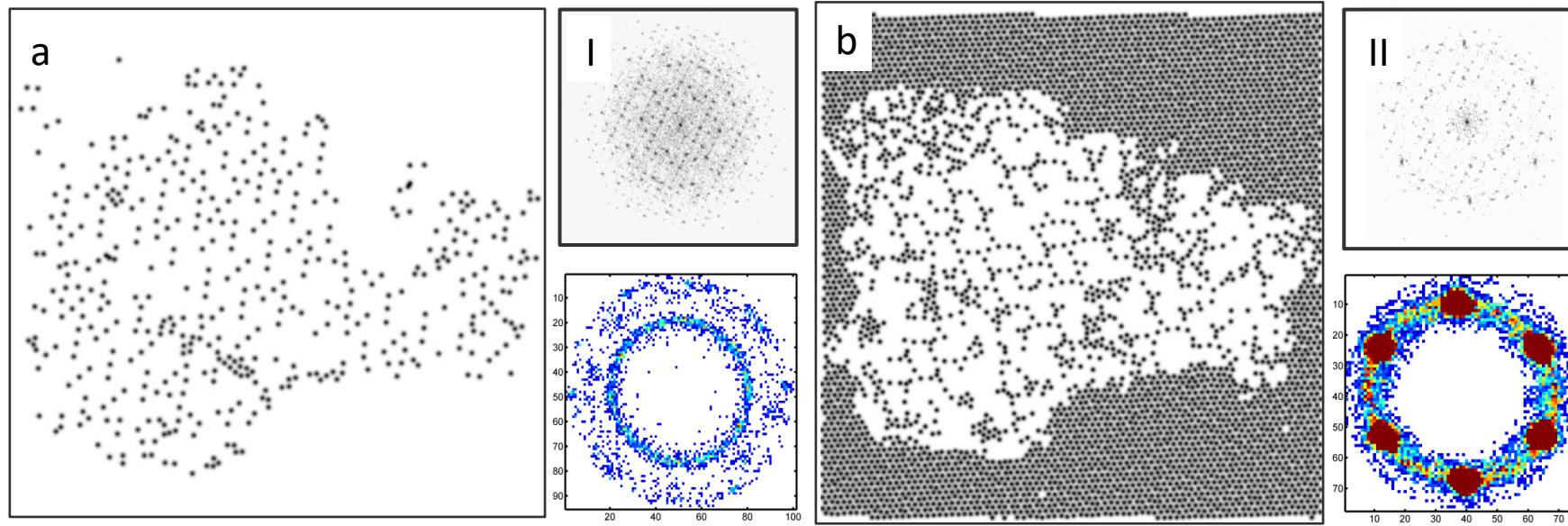
Local crystallography: k-means



A. BELIANINOV, Q. HE, M. KRAVCHENKO, S. JESSE, A. BORISEVICH, and S.V. KALININ, *Identification of phases, symmetries, and defects through local crystallography*, Nat. Comm. **6**, 7801 (2015).

Local crystallography

THE UNIVERSITY OF TENNESSEE  KNOXVILLE

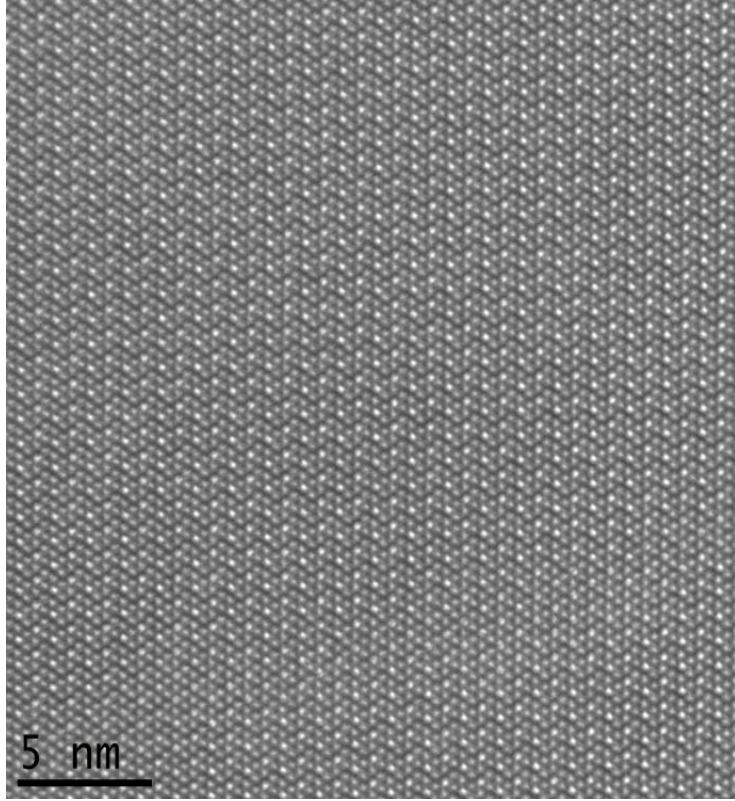


5 nm

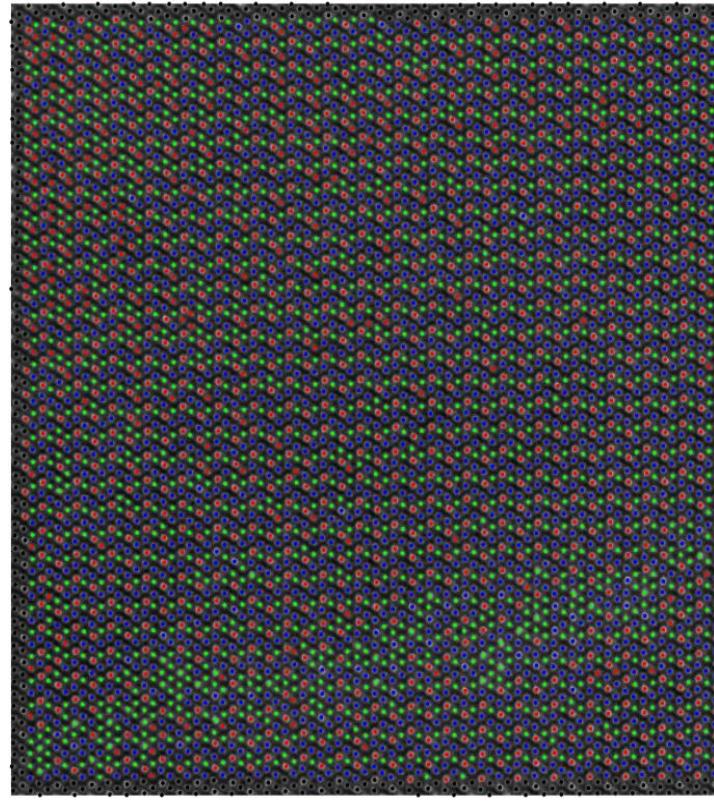
Local crystallography

THE UNIVERSITY OF TENNESSEE  KNOXVILLE

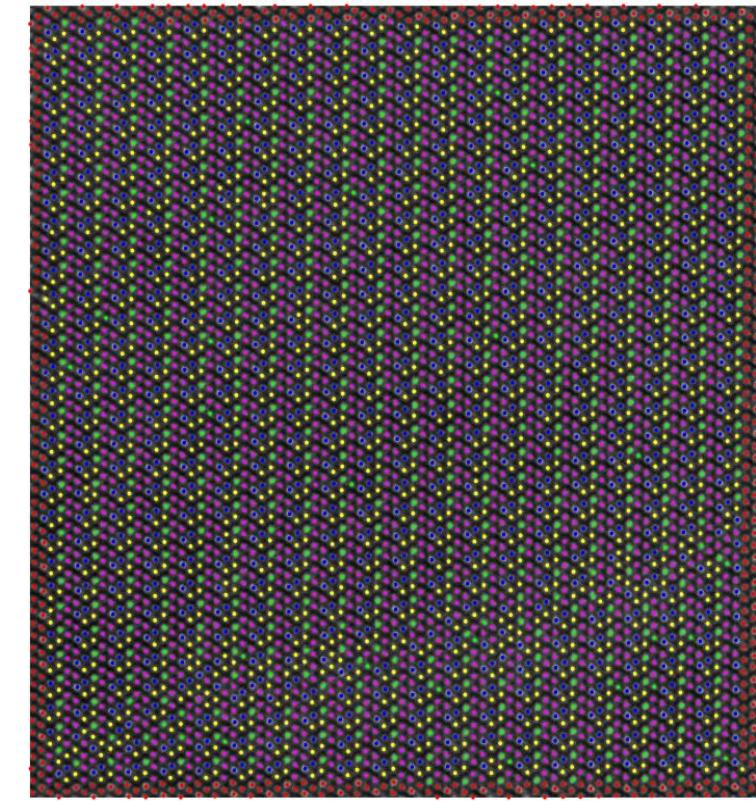
Image



K-means full vector



K-means angles



Normal modes: group theory

- **Group Representations** map group elements onto matrices, ensuring matrix multiplication aligns with the group operation.
- Molecules have **symmetries** defined by point groups linked to symmetry operations, such as rotations.
- Molecules have various **vibrational modes** with specific symmetries. Using the molecule's point group, one can deduce which modes are spectroscopically active.
- **Selection Rules:** Group representations dictate which vibrational modes appear in techniques like IR or Raman. Some modes may be IR-active but not Raman-active based on symmetry.
- **Vibrational frequencies** indicate energy differences between vibrational levels that are connected to group representations

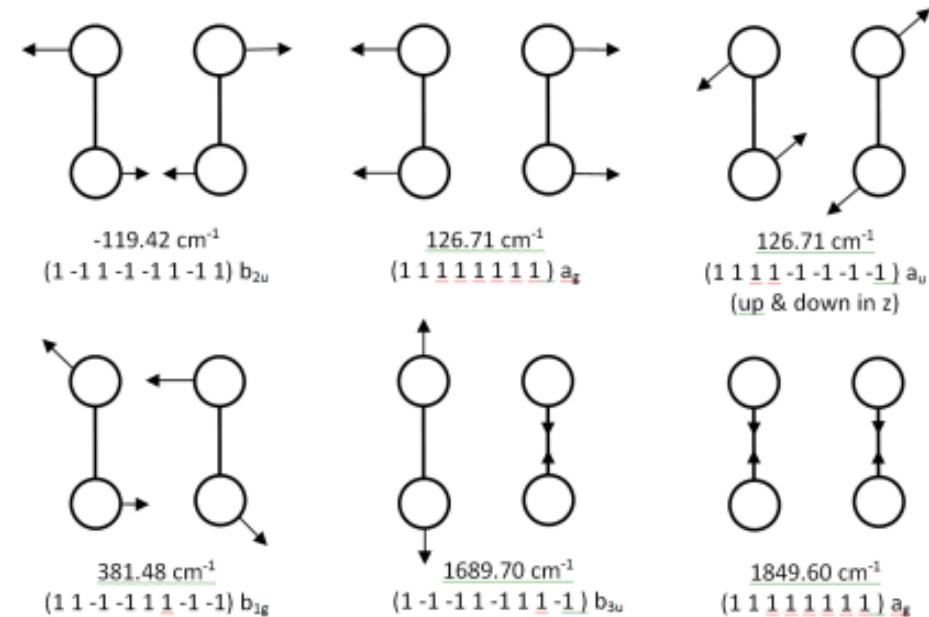
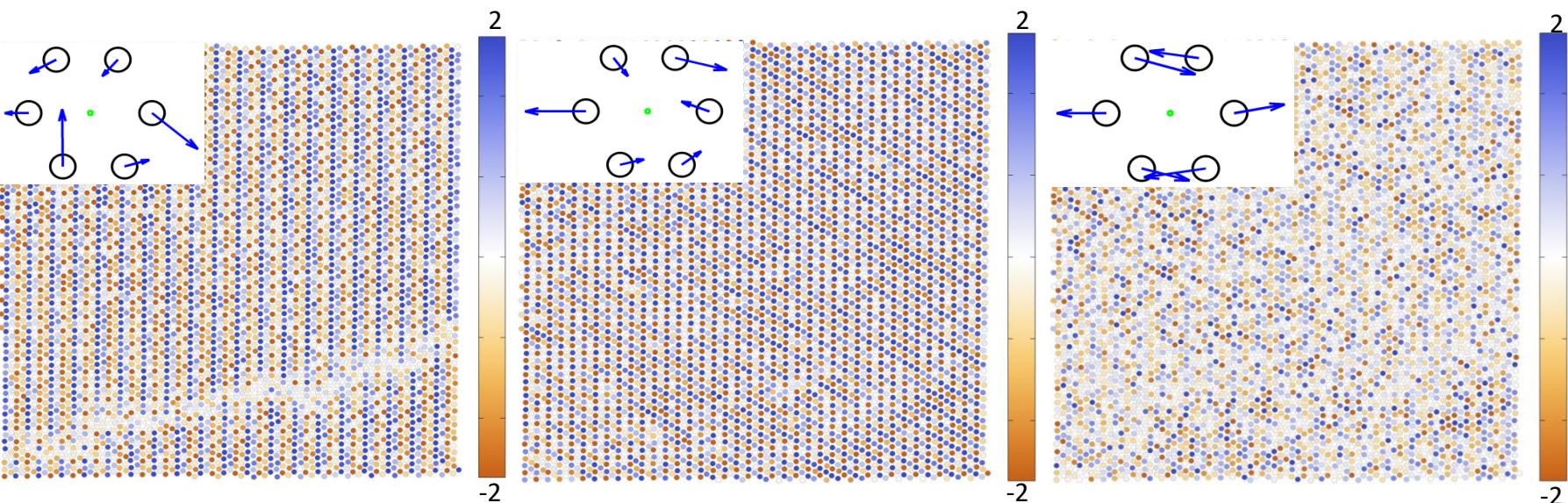
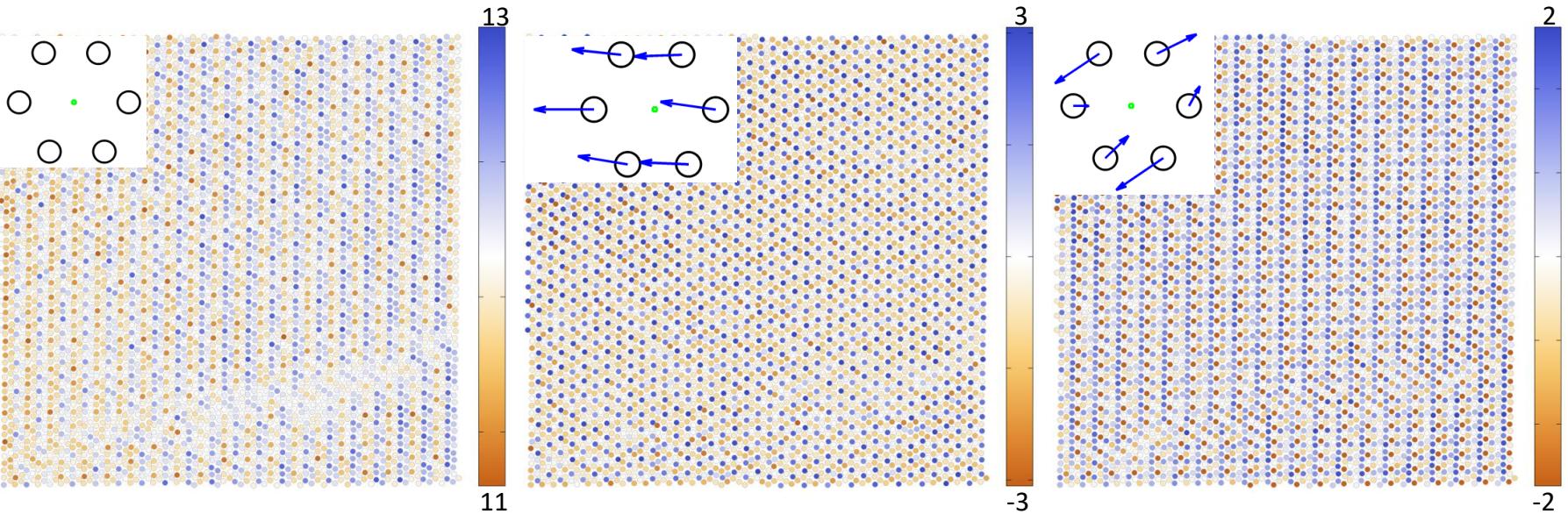


Figure 2: Sketches of Normal Modes of O_4^+

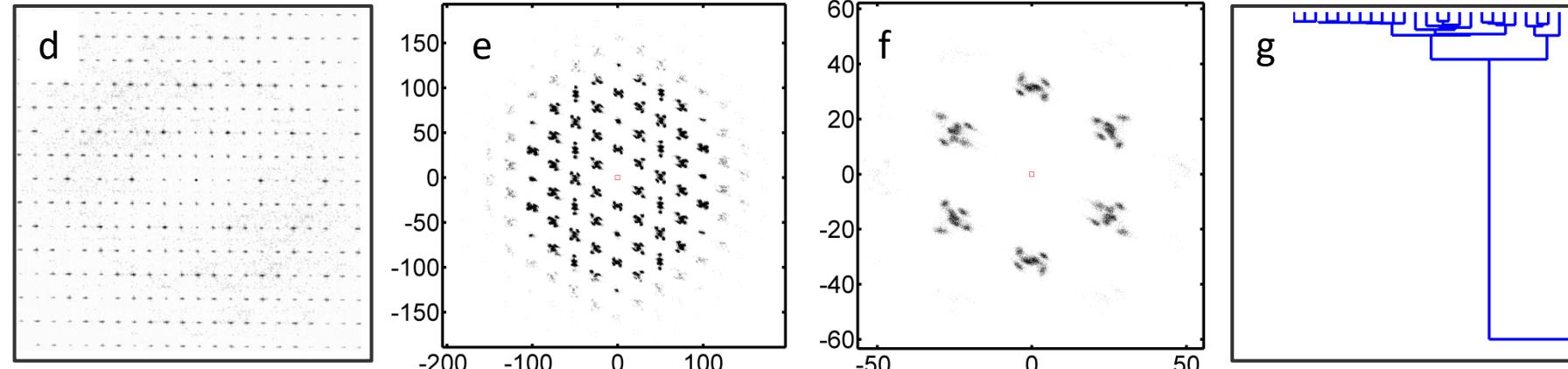
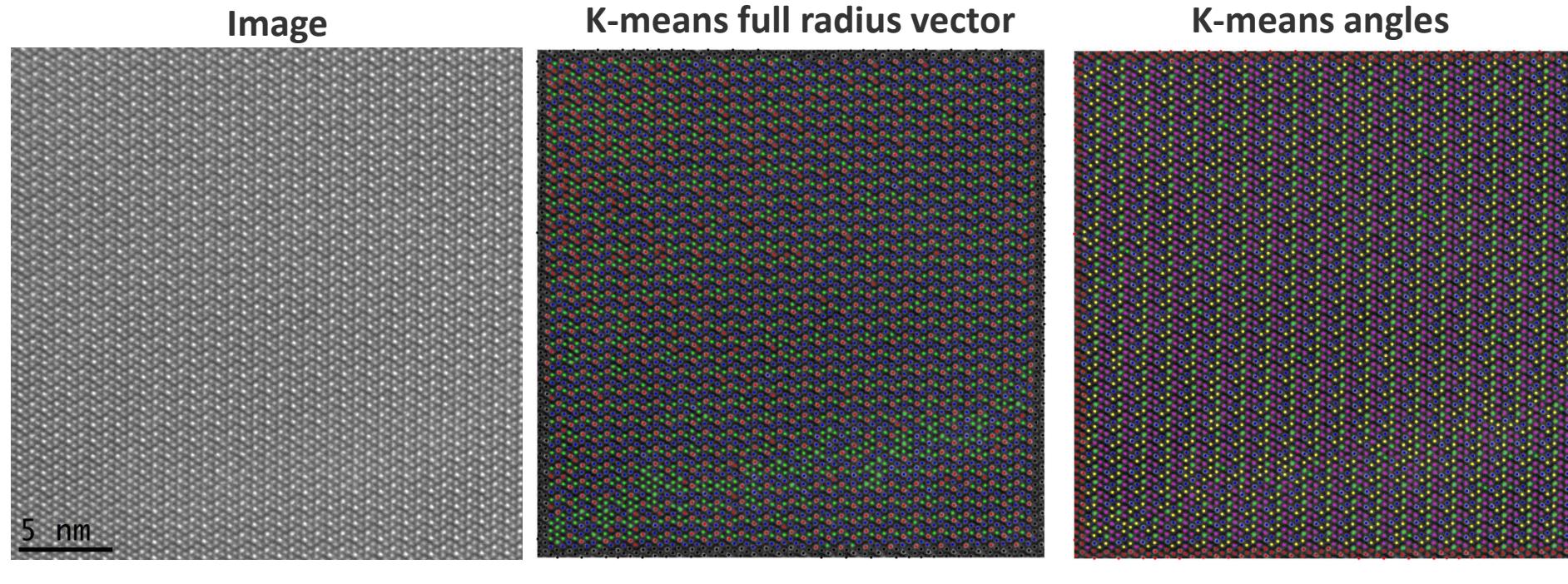
Table 2: Character Table for Point Group D_{2h}

D_{2h}	E	$C_2(z)$	$C_2(y)$	$C_2(x)$	i	$\sigma(xy)$	$\sigma(xz)$	$\sigma(yz)$		
A_g	1	1	1	1	1	1	1	1	x^2, y^2, z^2	
B_{1g}	1	1	-1	-1	1	1	-1	-1	R_z	xy
B_{2g}	1	-1	1	-1	1	-1	1	-1	R_y	xz
B_{3g}	1	-1	-1	1	1	-1	-1	1	R_x	yz
A_u	1	1	1	1	-1	-1	-1	-1		
B_{1u}	1	1	-1	-1	-1	-1	1	1	z	
B_{2u}	1	-1	1	-1	-1	1	-1	1	y	
B_{3u}	1	-1	-1	1	-1	1	1	-1	x	

Local crystallography: PCA

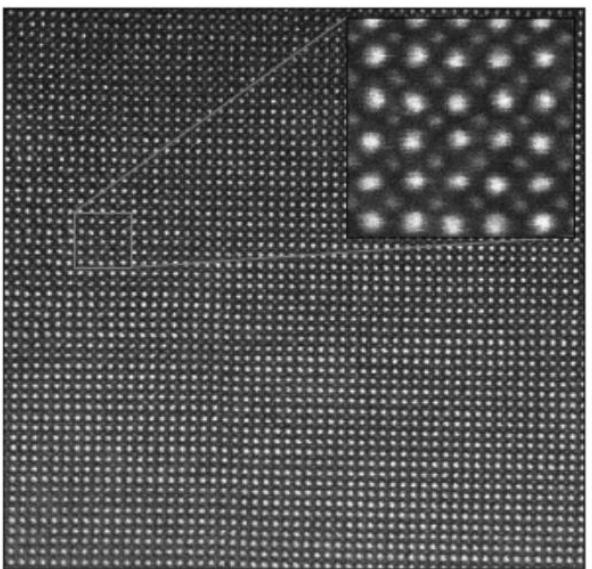


Local crystallography

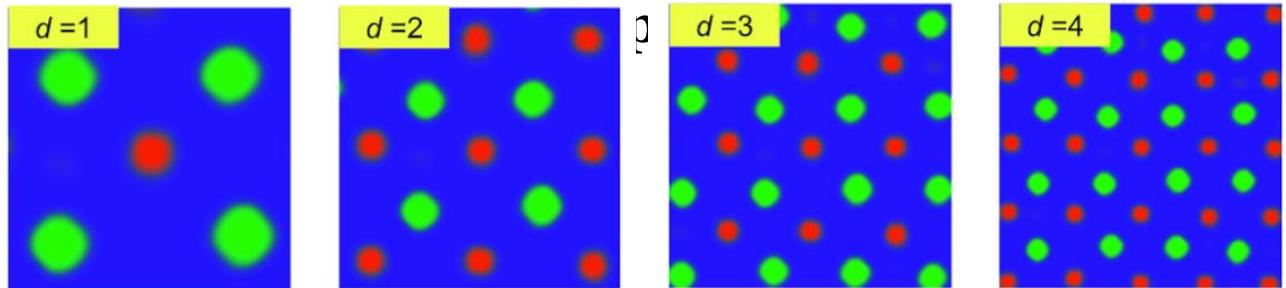


Local crystallography: FerroNET

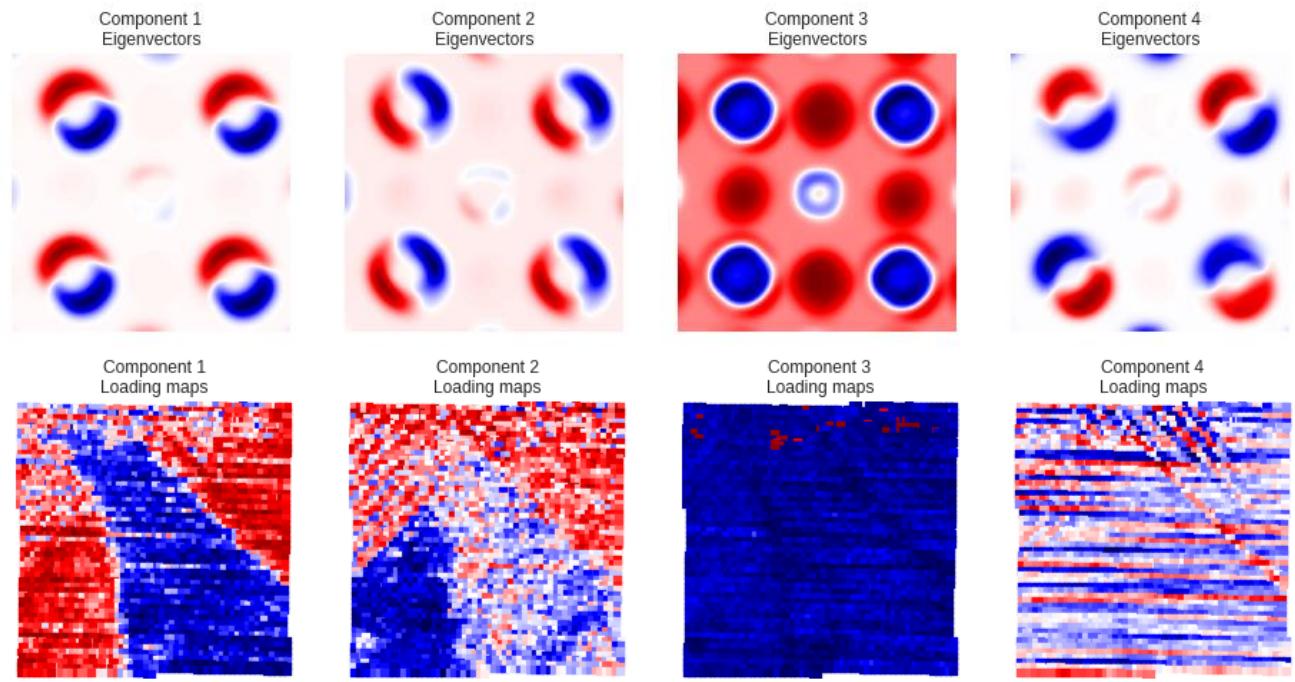
Experimental (LBFO film)



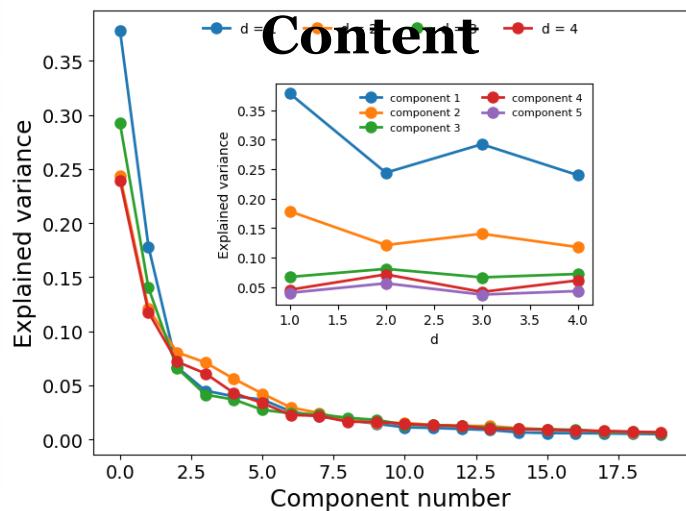
Building blocks (from neural network)



PCA eigenvectors and loading maps



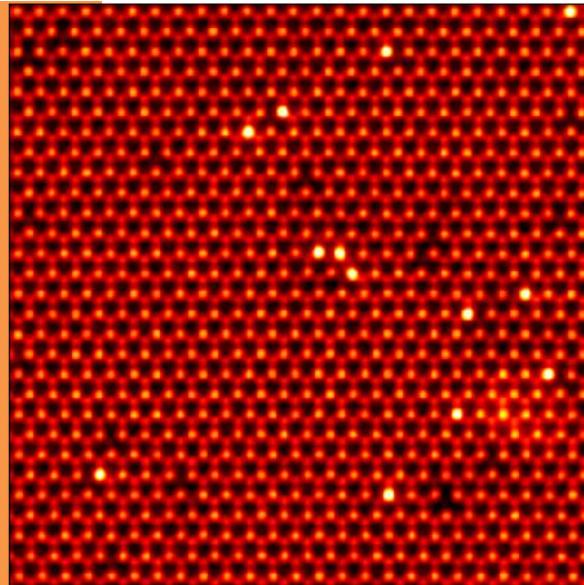
Information Content



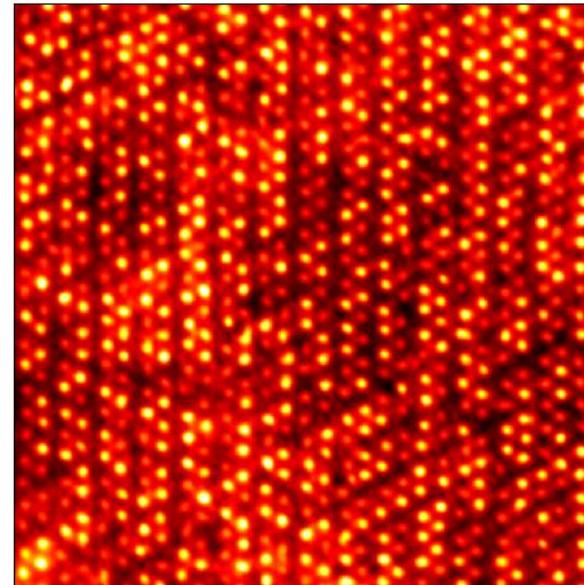
MoS₂-ReS₂

THE UNIVERSITY OF TENNESSEE  KNOXVILLE

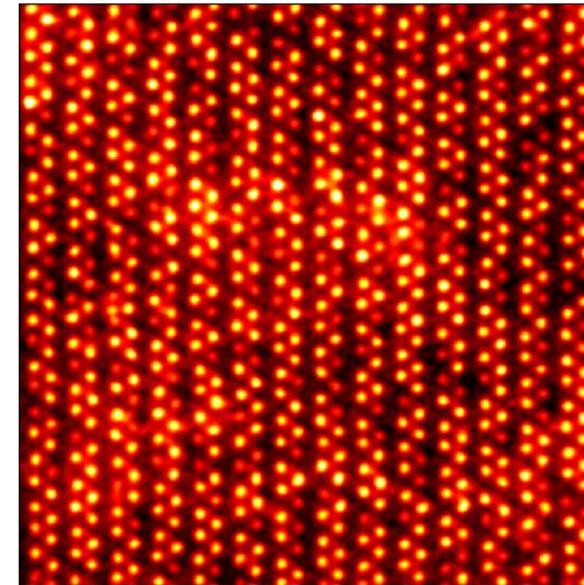
5% ReS₂



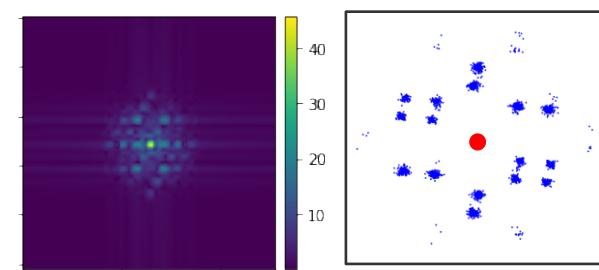
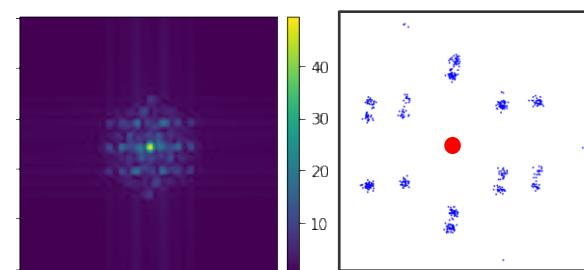
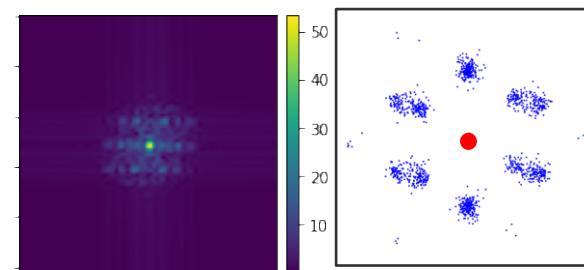
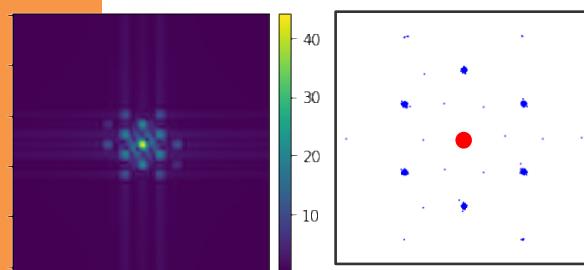
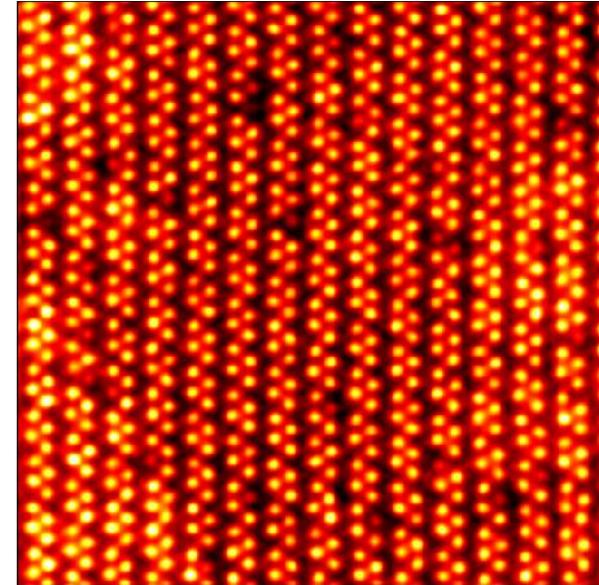
55% ReS₂



78% ReS₂

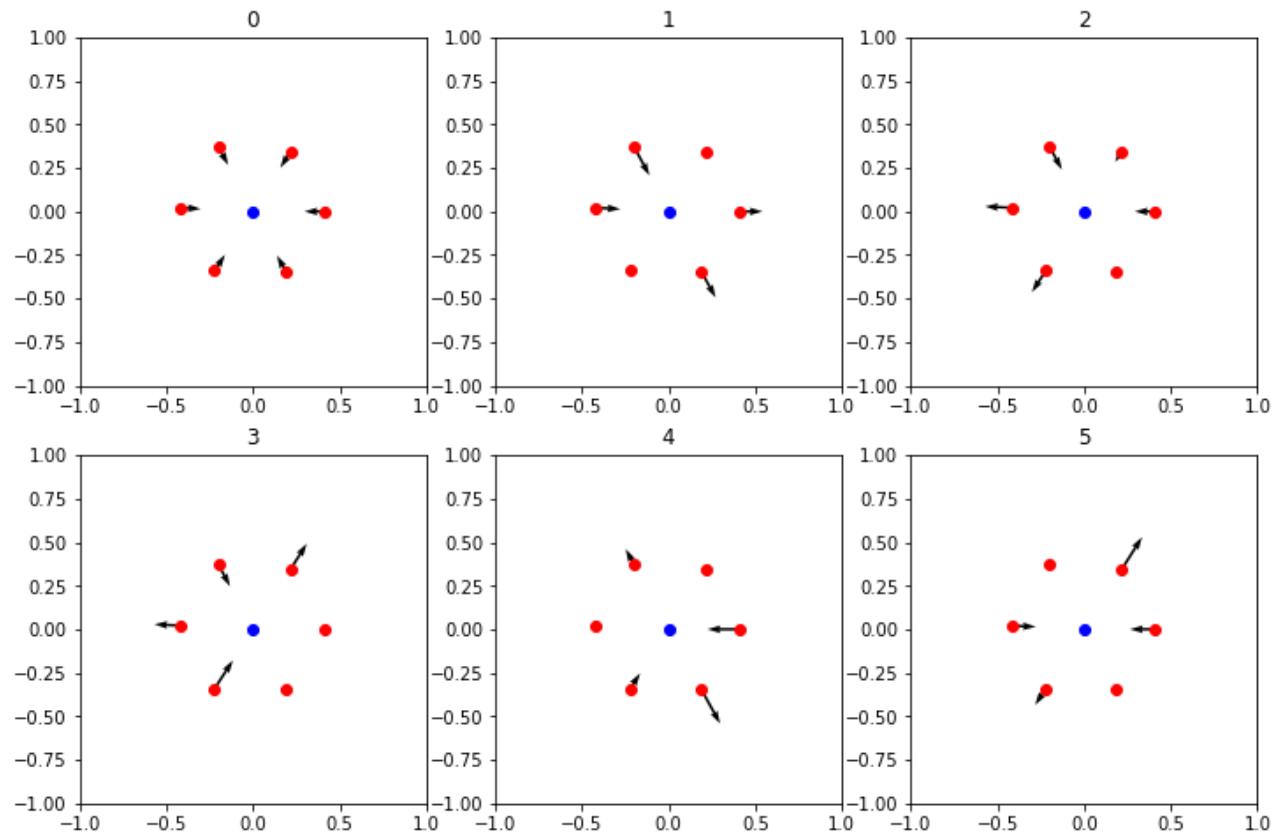
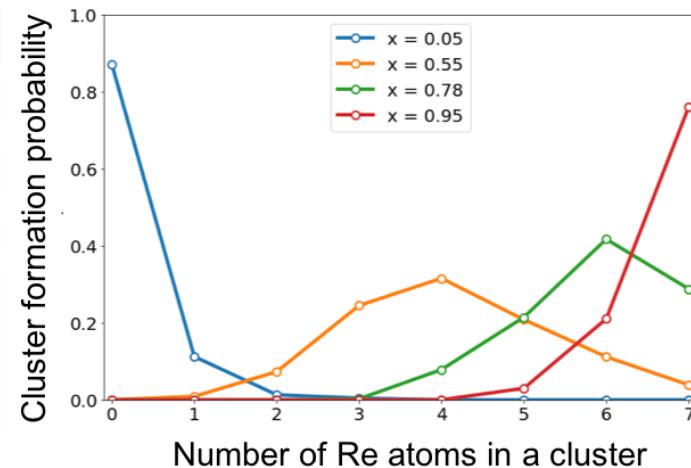
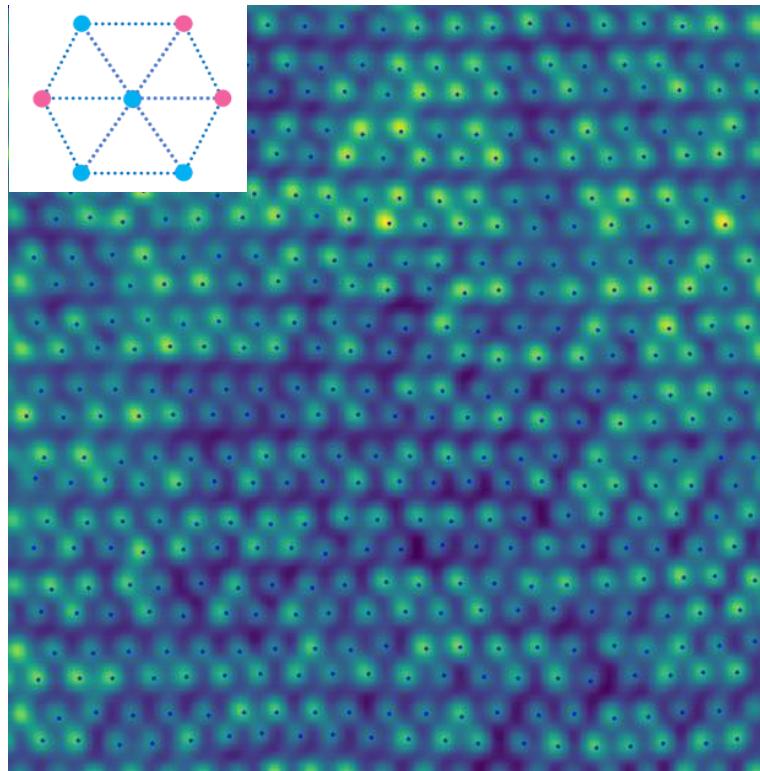


95% ReS₂



Data by Shize Yang and Matt Chisholm

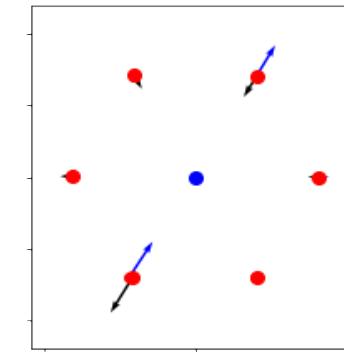
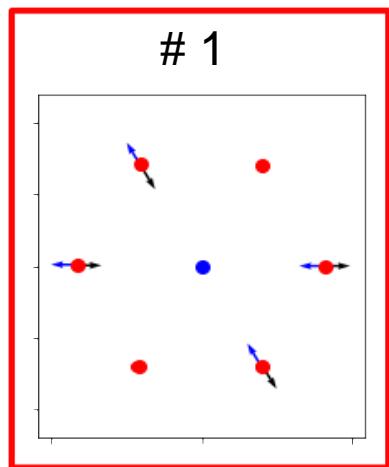
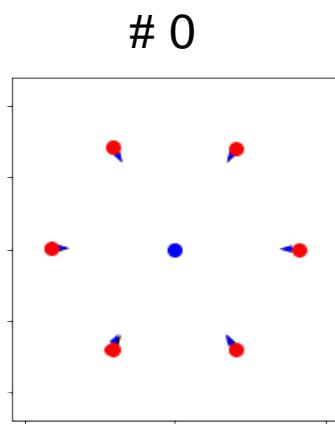
Local crystallography



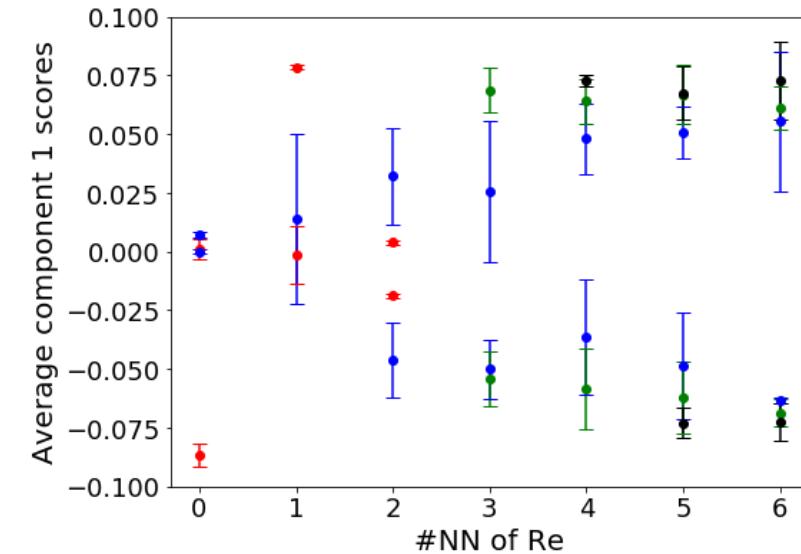
- Traditionally, the order parameter is defined based on symmetry and atomistic representation is established in the *ad hoc* manner
- But what if we define order parameter from the bottom up – based on the statistics of atomic distortions?
- And further correlate it to local chemical composition?

Describing the phase transition

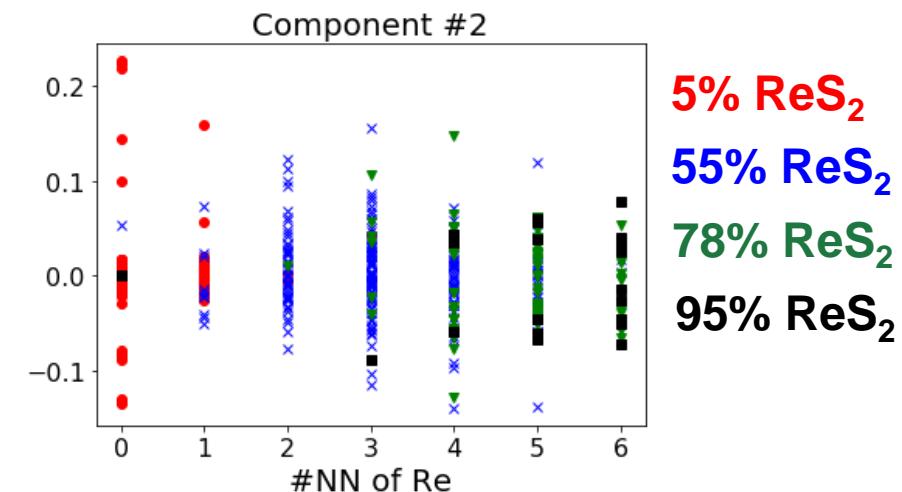
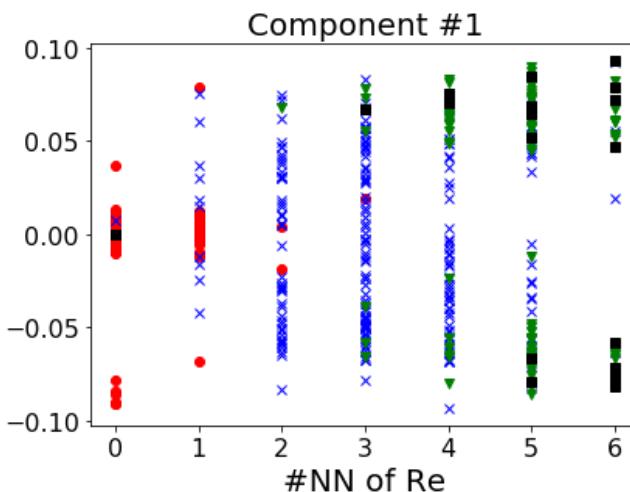
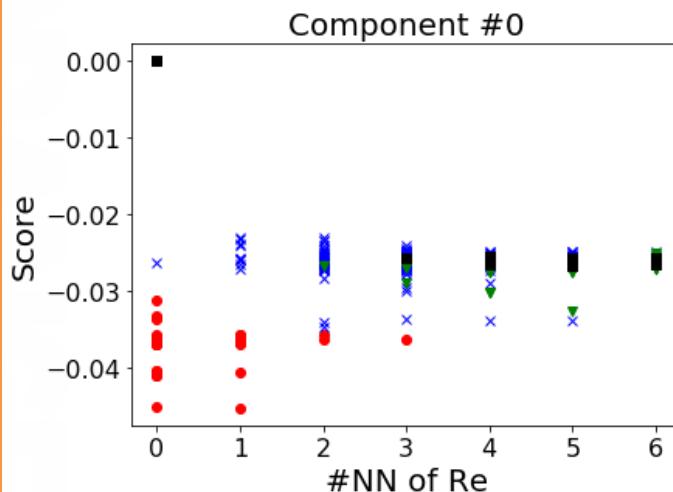
Three dominant distortion modes



Local symmetry breaking!



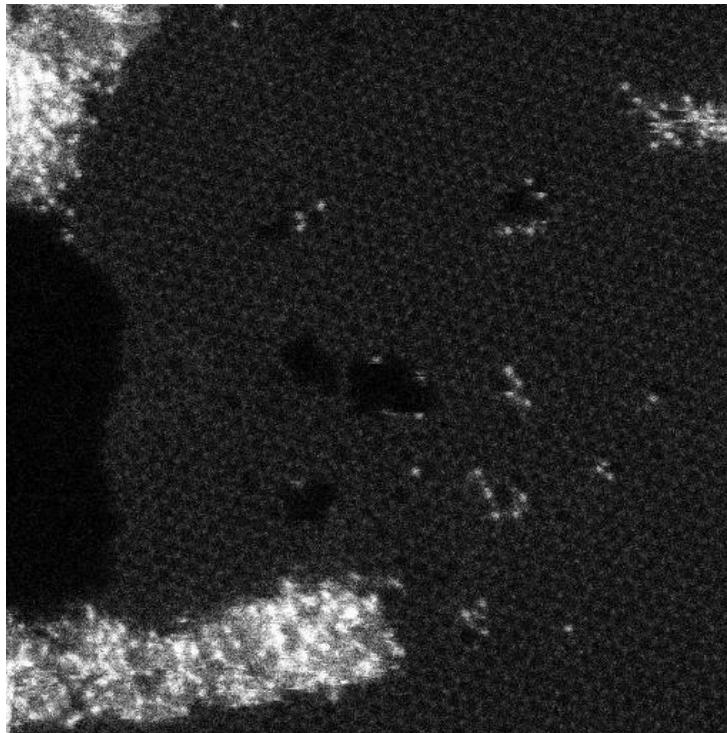
Mode distributions vs. global and local composition



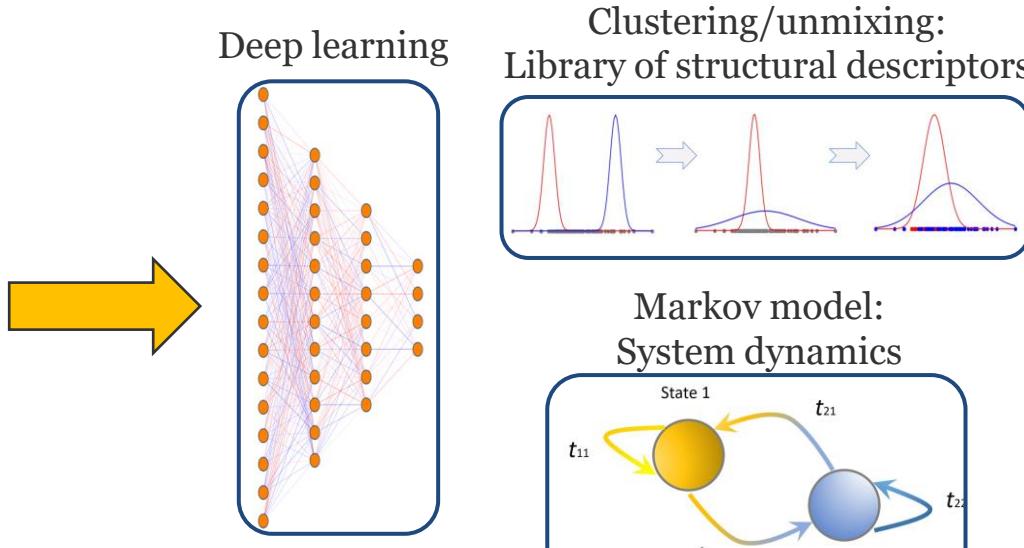
5% ReS₂
55% ReS₂
78% ReS₂
95% ReS₂

What about chemical dynamics?

Graphene+Si under e-beam



Deep & machine learning



Data collected by O. Dyck (ORNL)

1. Convert noisy experimental data into atomic positions/trajectories
→ **Deep convolutional neural networks**
2. Create libraries of structural descriptors
→ **Clustering/unmixing applied to the output of neural networks**
3. Analysis of dynamics and transition probabilities → **Markov modelling on the constructed classes**

Ziatdinov et al., ACS nano 11, 12742 (2017)

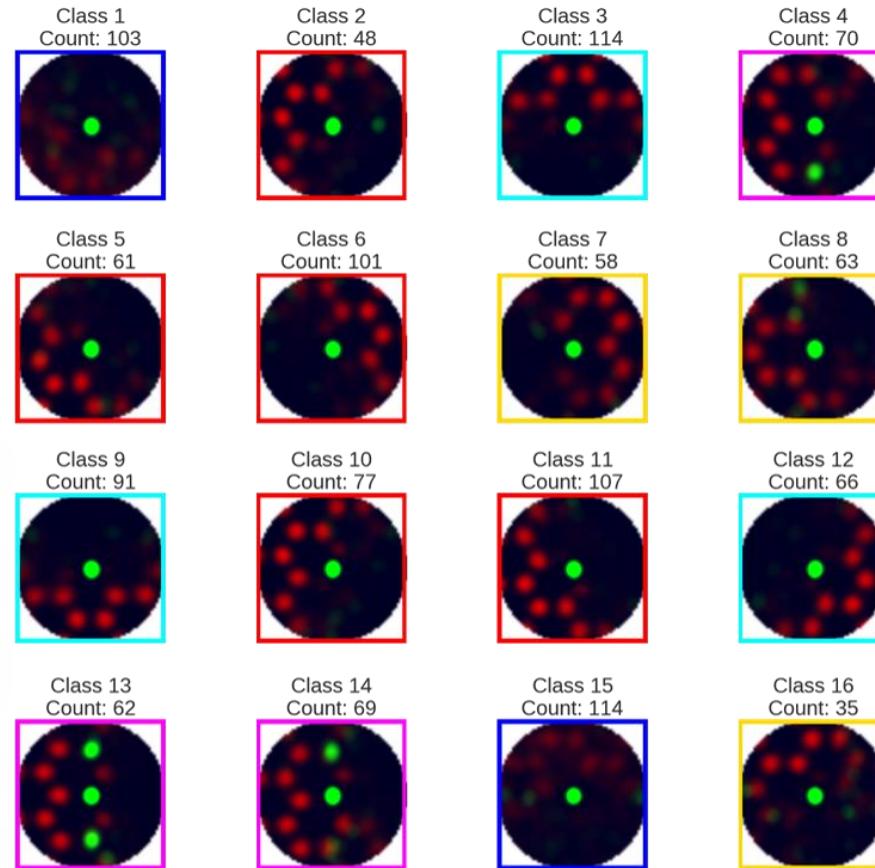
(2017)
Ziatdinov et al., Appl. Phys. Lett. 115, 052902 (2019)

Ziatdinov et al., npj Computational Materials 3, 31

Maksov et al., npj Computational Materials 5, 12 (2019)

Local crystallography

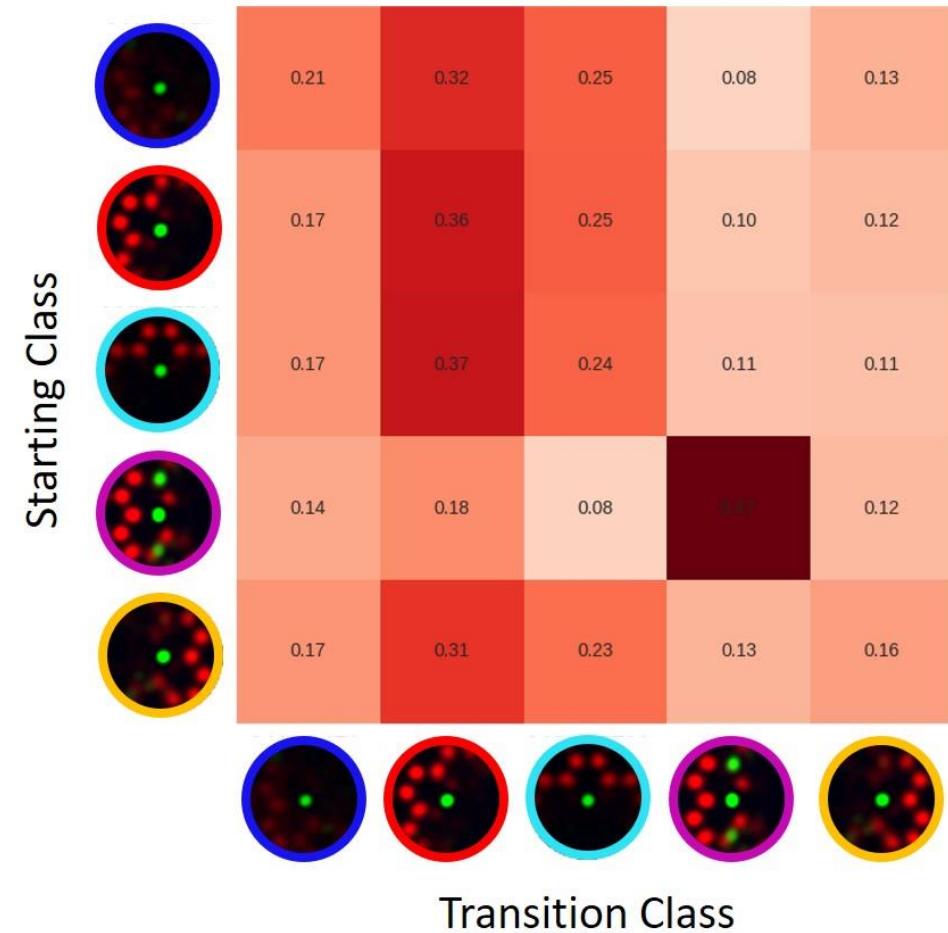
Derived classes of Si-C edge configurations



- Gaussian mixture model

- Discrete rotation symmetry + structural similarity algorithm

Transition probabilities matrix



- Markov state analysis