



Unsupervised topological learning approach of crystal nucleation in pure Tantalum

Sébastien Becker, Emilie Devijver, Rémi Molinier, Noël Jakse

► To cite this version:

Sébastien Becker, Emilie Devijver, Rémi Molinier, Noël Jakse. Unsupervised topological learning approach of crystal nucleation in pure Tantalum. Machine Learning and the Physical Sciences - Workshop at the 35th Conference on Neural Information Processing Systems (NeurIPS), Dec 2021, Vancouver, Canada. . hal-03559641

HAL Id: hal-03559641

<https://hal.science/hal-03559641>

Submitted on 7 Feb 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Unsupervised topological learning approach of crystal nucleation in pure Tantalum

Sébastien Becker (1, 2), Emilie Devijver (2), Rémi Molinier (3), and Noël Jakse (1)



(1) SIMaP, Univ. Grenoble Alpes, CNRS, Grenoble INP

(2) LIG, Univ. Grenoble Alpes, CNRS, Grenoble INP

(3) IF, Univ. Grenoble Alpes, CNRS



Abstract

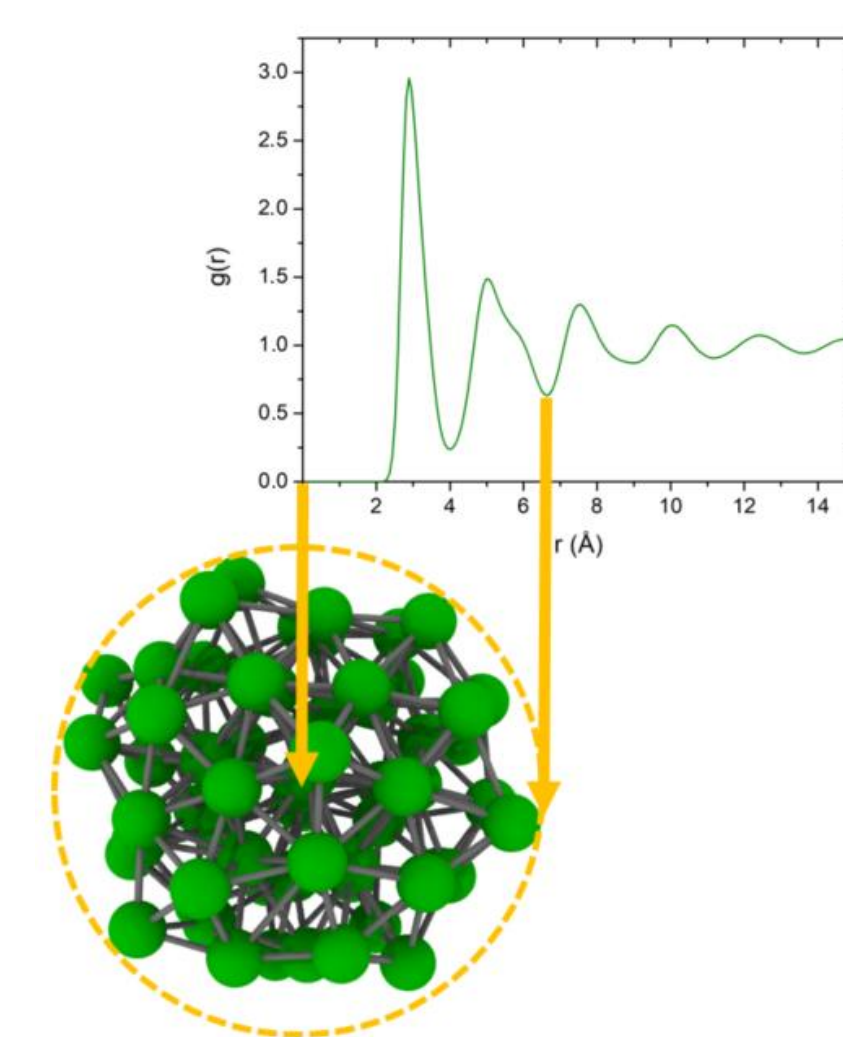
We propose an unsupervised machine learning method for the autonomous structural analysis of materials based on original topological descriptors of local atomic structures computed from persistent homology [1, 2]. From this protocol, a model is learned in course of nucleation in order to identify, without *a priori* on a system, clusters of atomic structures which arise in the process. This method has been applied to investigate the homogeneous nucleation of elemental Tantalum (Ta), on several molecular dynamics configurations of 10 million atoms at the nose of the time-temperature-transformation curve. A general behavior for all the nuclei (even the precritical ones) shows a concurrent emergence of the translational and orientational orderings in this pure metal.

Molecular dynamics simulations

Large-scale simulations with LAMMPS [3]:

- (N,P,T)
- $N = 10^7$
- $\Delta t = 2$ fs
- Verlet's algorithm
- EAM potential

Analysis on an isotherm at the nose of the time-temperature transformation curve



Train set of local atomic structures up to the second neighbor shell

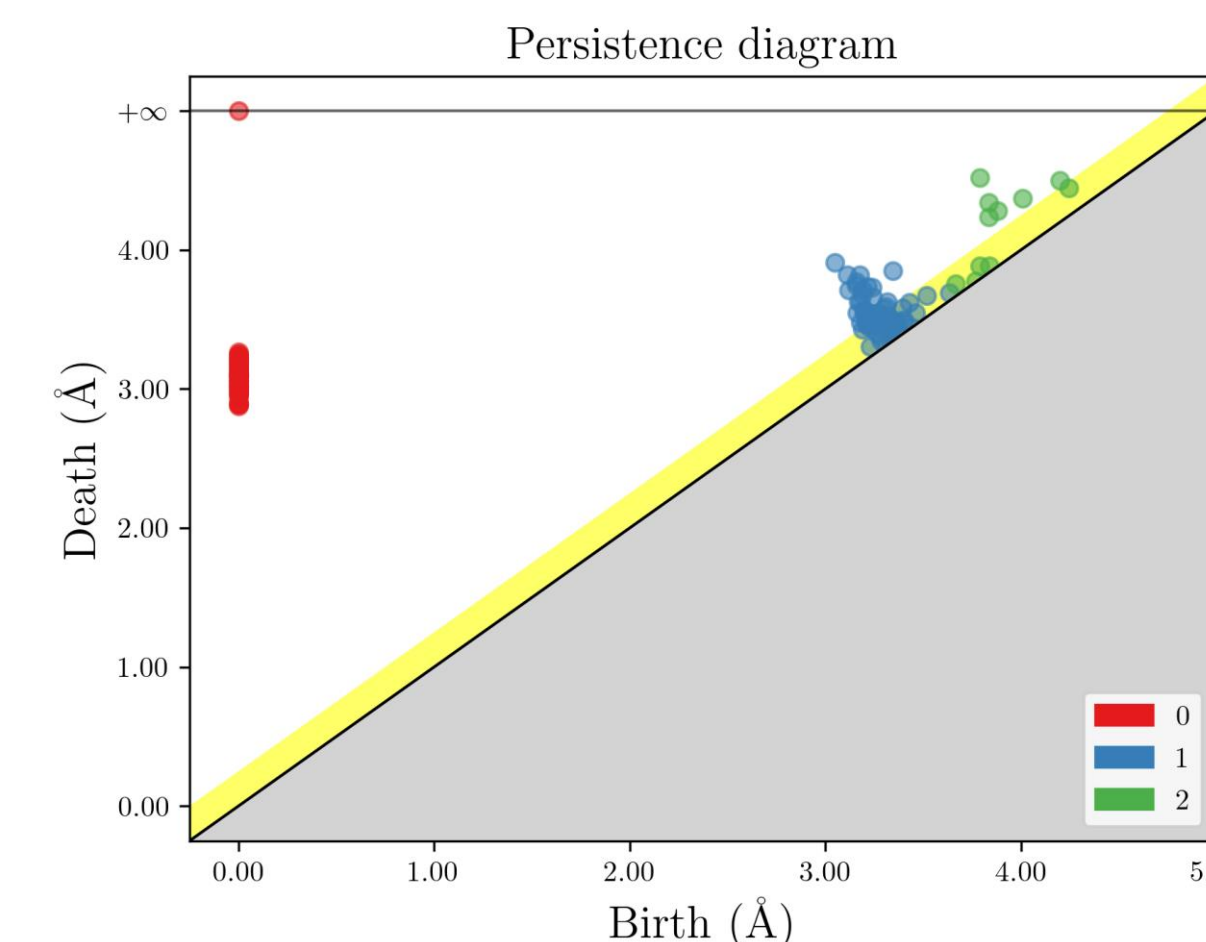
Model in course of nucleation

Inherent structures

Unsupervised learning approach

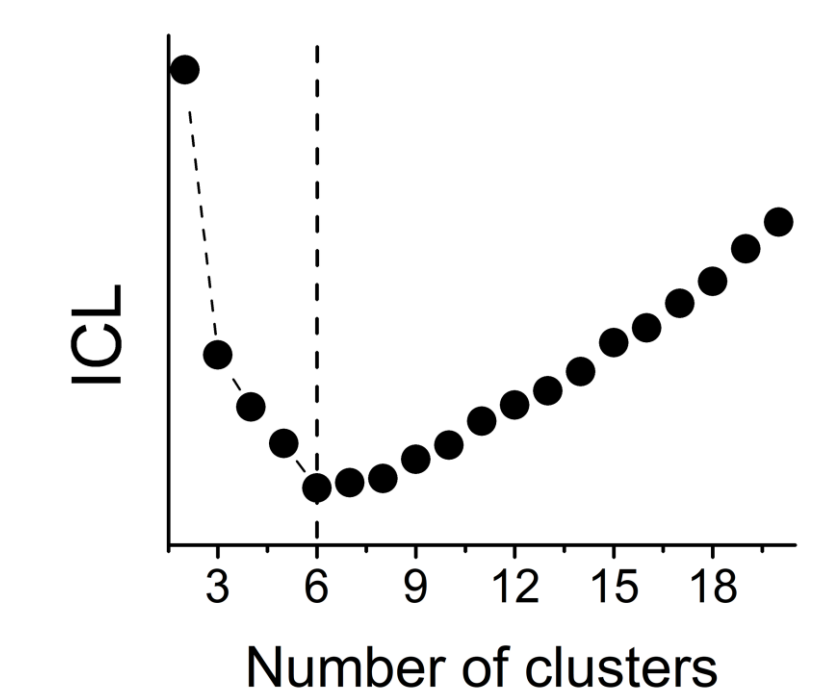
Encode local atomic structures using persistence diagrams from topological data analysis (TDA) [1, 2]

Subsampling approach to remove noises [4]



Topological vector [5] to construct a descriptor space of the train set

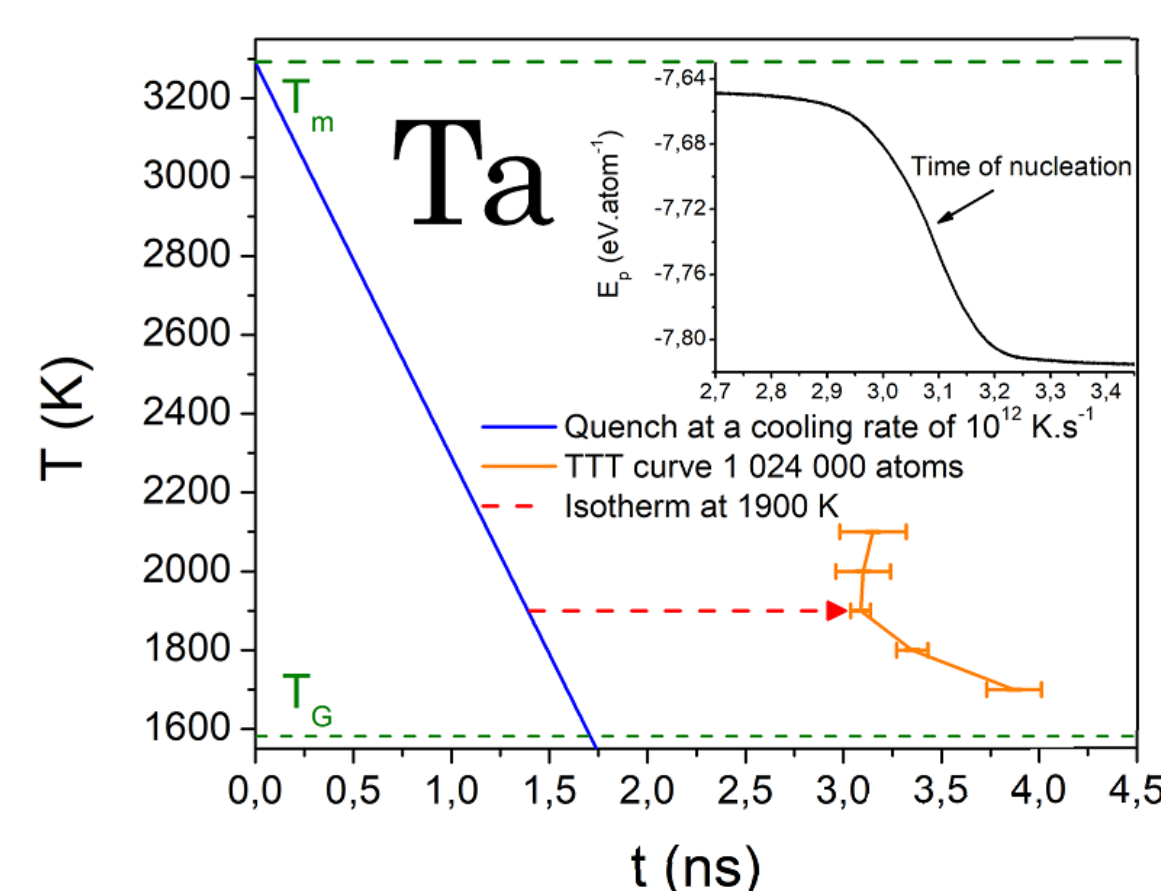
Gaussian mixture model (GMM) clustering with expectation-maximization algorithm [6]



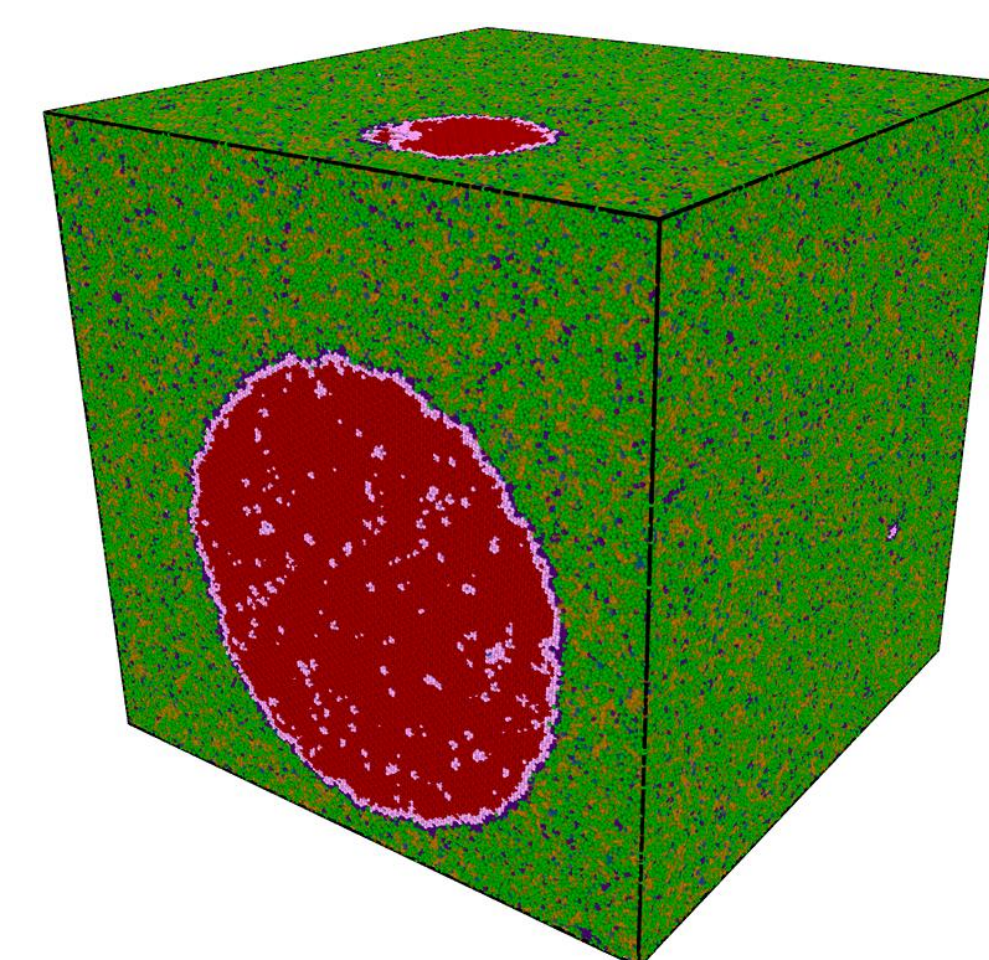
Integrated completed likelihood (ICL) [7] to initialize the number of Gaussian components (i.e. the number of clusters)

Autonomous description of the homogeneous crystal nucleation

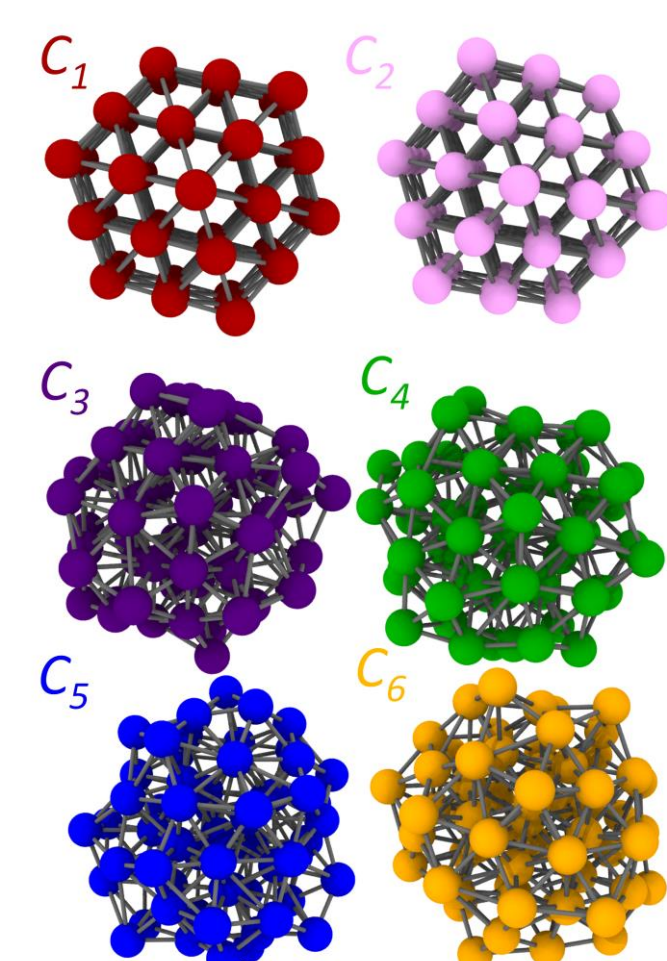
A. Building the model



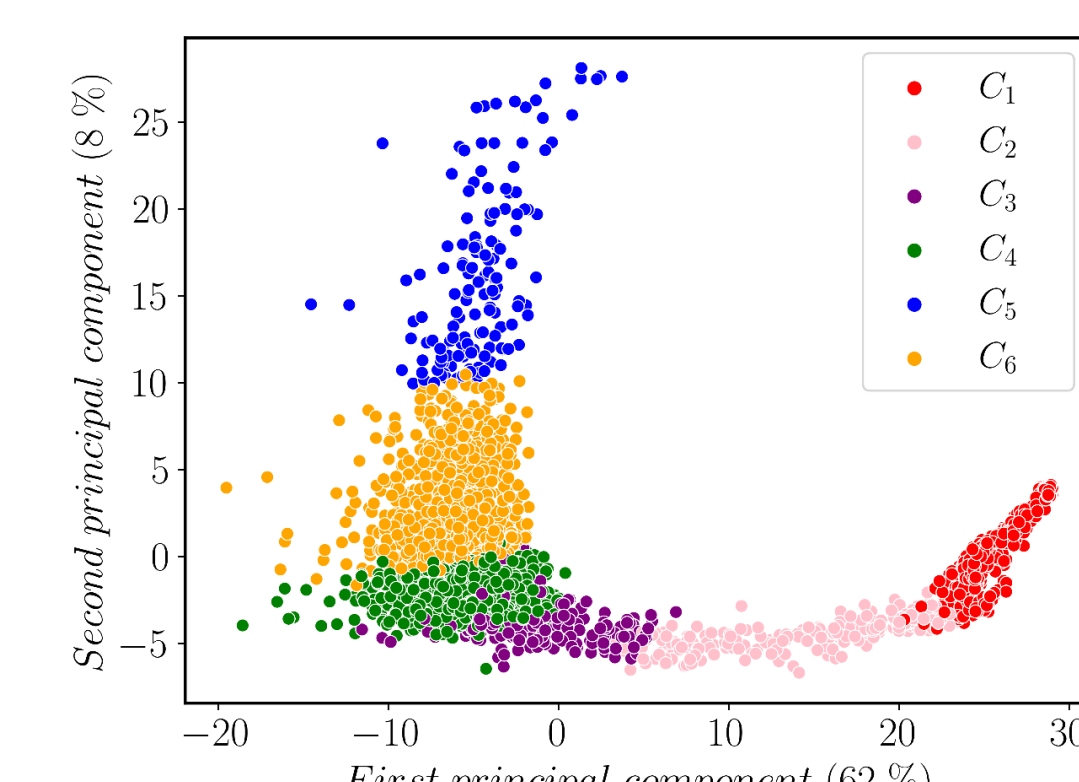
Isotherm analysis of crystal nucleation



Snapshot of the simulation with atoms coloured to according clusters

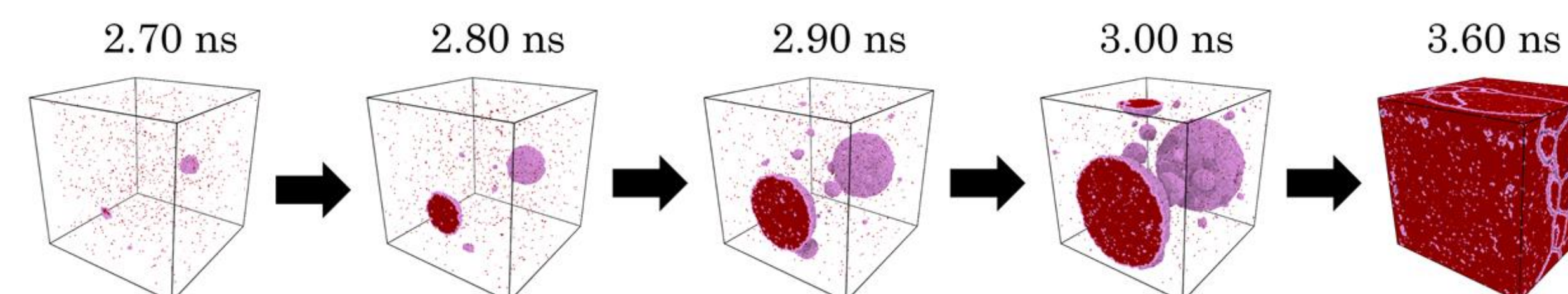


GMM clustering

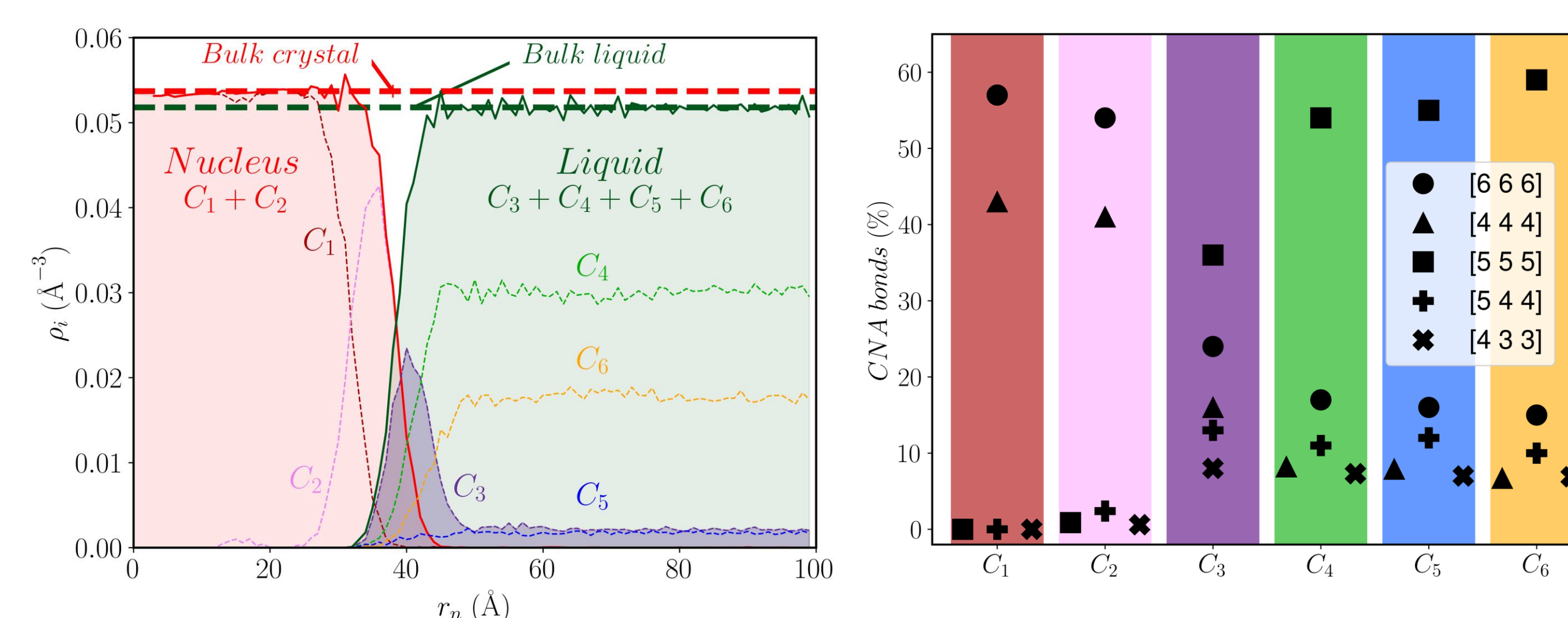


Principal component analysis of the clusters

B. Identification of the growing nuclei



C. Orderings analysis



Translational and orientational orderings arise simultaneously

Perspectives

This unsupervised TDA-GMM approach opens the way to autonomous application in the studies of structural dependant phenomena at the atomic scale. Extension of the nucleation analysis on multicomponent systems is especially relevant for a deeper inspection and control of such mechanisms to enhance materials design. As a perspective, it would be interesting to extend the method to learn the time evolution, e.g., through recent generalization of the persistent homology to time series [8].

References

- [1] H. Edelsbrunner, D. Letscher, and A. Zomorodian, Discrete Comput Geom **28**, 511 (2002).
- [2] A. Zomorodian and G. Carlsson, Discrete Comput Geom **33**, 249 (2005).
- [3] S. J. Plimpton, J. Comput. Phys. **117**, 1 (1995).
- [4] B. T. Fasy, F. Lecci, A. Rinaldo, L. Wasserman, S. Balakrishnan, and A. Singh, Ann. Statist. **42**, (2014).
- [5] M. Carriere, S. Y. Oudot, and M. Ovsjanikov, Computer Graphics Forum **34**, 1 (2015).
- [6] A. P. Dempster, N. M. Laird, and D. B. Rubin, Journal of the Royal Statistical Society: Series B (Methodological) **39**, 1 (1977).
- [7] C. Biernacki, G. Celeux, and G. Govaert, IEEE Trans. Pattern Anal. Machine Intell. **22**, 719 (2000).
- [8] N. Ravishanker, and R. Chen, WIREs Computational Statistics **13**, 3, e1548 (2021).



Fourth Workshop on Machine Learning and the Physical Sciences (NeurIPS 2021)