Density-Aware Phase Segmentation in Sparse 3D Atomistic Data

ABIN SHAKYA and BIJAYA B. KARKI*, Louisiana State University, USA

Segmenting 3D atomistic simulation data is challenging due to sparsity, irregular structure, and the absence of clear phase boundaries. Unlike conventional segmentation tasks, phase identification in molecular systems depends on subtle density variations rather than geometric cues. We address this problem in systems exhibiting Fe-rich (metallic) and Fe-poor (silicate) phases with gradual compositional transitions. Traditional voxel-based methods often fail in small supercells due to noise and fragmentation. We propose a structure-aware segmentation pipeline that combines density smoothing, unsupervised clustering, and periodicity-aware connected component analysis. Our method achieves consistent, physically meaningful segmentation across time, enabling robust compositional analysis and trace element partitioning in sparse, constrained data.

CCS Concepts: • **Computing methodologies** → **Unsupervised learning**; • **Applied computing** → *Phase Segmentation.*

Additional Key Words and Phrases: Phase Characterization, Molecular Dynamics, Periodic Boundary Condition

1 Introduction

Understanding phase separation in atomistic systems is critical for interpreting molecular simulations of complex materials. We study a 520-atom Fe–Mg–Si–O–N system ($Fe_{85}Mg_{104}Si_{72}O_{251}N_8$), representative of high-pressure planetary interiors, where two compositional phases emerge: an Fe-rich (metallic) and an Fe-poor (silicate) region. This setup models early Earth conditions, where metal–silicate separation drives core–mantle differentiation. Our goal is to segment these phases and quantify trace element behavior—particularly nitrogen—using the nitrogen partitioning coefficient (K_D^{N}), defined as the ratio of nitrogen's weight percent in Fe-rich to Fe-poor regions.

Phase identification in this setting is complicated by sparse and irregular atomic arrangements and the lack of sharply defined boundaries between regions. Unlike conventional segmentation problems [3, 6], here the distinction must be made based on subtle differences in local density rather than geometric or visual cues. Traditional methods such as geometric fitting or threshold-based binning [7] often fail due to noise, limited resolution, and fragmented spatial continuity.

To address these challenges, we introduce a structure-aware segmentation pipeline that integrates smoothed Fe density estimation, unsupervised clustering, and periodic-aware connected component analysis to extract contiguous compositional phases. This enables robust and consistent phase identification, facilitating reliable compositional analysis across time-resolved simulation frames.

2 Related Work

Traditional geometric methods like convex hulls [1] and alpha shapes [2] are commonly used to delineate spatial regions but perform poorly in sparse, periodic atomistic simulations. Convex hulls tend to overestimate boundaries in non-convex or hollow structures, while alpha shapes require fine parameter tuning and often miss meaningful features in irregular data. Both approaches ignore periodic boundary conditions, resulting in discontinuities. Voxel-based binning, used in our previous work [7], also suffers from noise and resolution issues in small supercells.

To address these challenges, we introduce a structure-aware segmentation pipeline that transforms atomic positions into a smoothed Fe density field, applies unsupervised clustering, and incorporates periodicity-aware connectivity filtering. This enables accurate and robust phase segmentation in sparse 3D atomic systems.

3 Approach

We propose a structure-aware segmentation pipeline for identifying compositional phases (e.g., Fe-rich and Fe-poor) from sparse 3D atomistic simulation frames. The method is fully unsupervised, physically informed, and designed to operate under periodic boundary conditions (PBC). The key steps are as follows:

3.1 Binning with PBC-aware Tiling

To estimate the spatial distribution of Fe atoms, we discretize atomic positions into a 3D histogram grid (100×100×100 bins). To respect PBC, we tile each simulation frame by virtually replicating atomic coordinates across ±1 box length in all three dimensions. This ensures density continuity at simulation box edges.

3.2 Gaussian Smoothing

Because the underlying atomic data is sparse and irregularly sampled, we apply a Gaussian filter to the binned histogram. This produces a continuous density field, helping suppress noise while preserving compositional gradients.

3.3 KMeans Clustering

We flatten the smoothed density field and apply KMeans clustering (K=2) to distinguish high-density (Fe-rich) and low-density (Fe-poor) regions. Unlike thresholding-based methods, this data-driven approach adapts to local density distributions across different frames.

3.4 Connectivity Filtering via Union-Find

We enforce spatial continuity of the Fe-rich region using a union-find algorithm that tracks face-connected components while accounting for Periodic Boundary Condition (PBC). We retain the largest connected Fe-rich cluster and remove small fragmented patches from the Fe-poor region.

3.5 Boundary Region Assignment

To capture the diffuse nature of phase transitions, we define a boundary region between Fe-rich and Fe-poor zones using symmetric distance transforms. Voxels within a specified distance (0.17 Å in our case) from both Fe-rich and Fe-poor regions are labeled as boundary. This allows us to isolate interfacial regions where compositional mixing or phase exchange is most likely to occur.

Authors' Contact Information: Abin Shakya, ashaky3@lsu.edu; Bijaya B. Karki, bbkarki@lsu.edu, Louisiana State University, Baton Rouge, Louisiana, USA.

3.6 Atom Classification and Composition Analysis

Voxel-level labels (Fe-rich, Fe-poor, boundary) are assigned to individual atoms based on their bin indices. For each region, we calculate element-wise counts and weight percentages using known atomic masses. This analysis is performed across all simulation frames, and the results are aggregated to assess segregated phase composition over time.

4 Results

We applied our segmentation pipeline to a 520-atom Fe–Mg–Si–O–N simulation, initialized with Bulk Earth composition and evolved under high-pressure conditions. The analyzed snapshots capture a phase-separated state in which heavier elements (e.g., Fe) and lighter elements (e.g., Mg, Si, O) have spatially segregated into two distinct regions—mimicking the metal–silicate separation that underlies core–mantle differentiation.

Figure 1 shows the atomic configuration of a representative snapshot. Although the identity of each atom is known, the underlying phases are not visually separable, motivating the need for automated segmentation.

Figure 2 illustrates the segmentation output across central XY, YZ, and ZX slices. While segmentation is performed in 3D, these 2D views support interpretation. The top and middle rows display Fe-rich (yellow), Fe-poor (black), and fuzzy boundary (red) regions without and with a 0.34 Å buffer. The bottom row shows the smoothed Fe density input used for clustering.

We analyzed 25 uniformly sampled snapshots and computed the average weight percent of each element in Fe-rich and Fe-poor regions. Table 1 summarizes the results. The Fe-rich region exhibits >96% Fe content with minimal light elements, closely resembling Earth's core, while the Fe-poor region contains high concentrations of O, Mg, and Si—consistent with mantle-like composition. This supports the physical validity of our segmentation approach.

We further quantified nitrogen partitioning using the coefficient:

$$K_D^{\mathrm{N}} = \frac{\text{wt\% of N in Fe-rich}}{\text{wt\% of N in Fe-poor}}.$$

We find $K_D^{\mathrm{N}} \approx 9.0$ without boundary buffering and $K_D^{\mathrm{N}} \approx 11.7$ with buffering, confirming nitrogen's strong preference for the Fe-rich phase.

Table 1. Elemental weight percent (%) in segmented regions (avg. over 25 snapshots) vs. Bulk Earth model [4, 5].

Elem	**w/ Bound.**		**No Bound.**		**Core**	**Mantle**
	Fe-rich	Fe-poor	Fe-rich	Fe-poor	Model	Model
Fe	96.5	0.11	94.7	0.11	95.5	6.9
Mg	0.00	30.6	0.00	30.0	–	28.0
Si	0.86	23.1	1.41	23.2	3.0	21.0
O	0.96	46.1	1.99	46.5	1.5	44.0
N	1.64	0.14	1.89	0.21	–	–

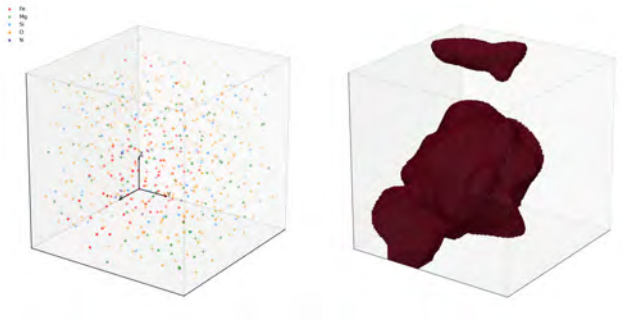

Fig. 1. 3D atomic configuration and Fe-rich phase segmentation from a representative molecular dynamics snapshot. Left: Atomic positions are visualized with atoms color-coded by element — Fe (red), Mg (green), Si (blue), O (orange), and N (violet). Right: Segmented Fe-rich regions are shown as translucent crimson voxel volumes, highlighting spatially clustered metal-rich domains within the simulation box.

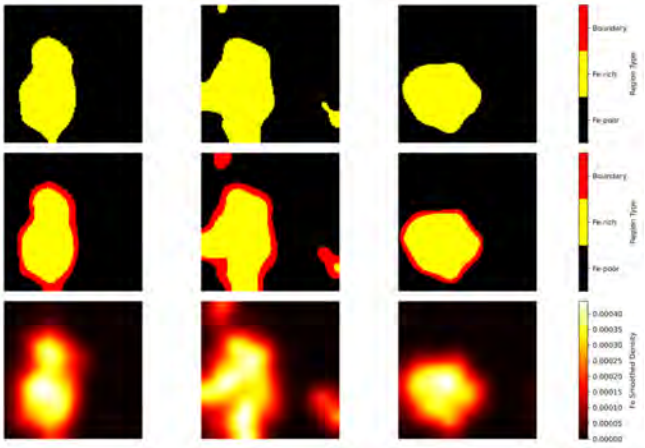

Fig. 2. Phase segmentation and Fe density visualization across central slices (XY, YZ, ZX). **Top:** Segmentation without boundary. **Middle:** Segmentation with 0.34 Å fuzzy boundary (red). **Bottom:** Smoothed Fe density used for clustering.

5 Discussion and Conclusion

We present a structure-aware segmentation pipeline for identifying segregated phases in sparse 3D atomistic data. By combining density smoothing, unsupervised clustering, and periodicity-aware connected component filtering—including the removal of small, disconnected patches—the pipeline addresses key limitations of traditional geometric methods. The resulting Fe-rich and Fe-poor phase assignments reproduce expected compositional trends, including nitrogen enrichment in the Fe-rich phase. These results demonstrate that accurate and physically meaningful segmentation is achievable even in sparse, irregular systems, enabling robust analysis of phase behavior and trace element partitioning in complex materials.

References

[1] Bernard Chazelle. 1993. An optimal convex hull algorithm in any fixed dimension. *Discrete & Computational Geometry* 10, 4 (1993), 377–409.

[2] Herbert Edelsbrunner and Ernst P. Mücke. 1994. Three-dimensional alpha shapes. *ACM Transactions on Graphics (TOG)* 13, 1 (1994), 43–72. doi:10.1145/174462.156635

[3] Truc Le and Ye Duan. 2018. PointGrid: A Deep Network for 3D Shape Understanding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 9204–9214. doi:10.1109/CVPR.2018.00959

[4] William F. McDonough. 2003. Compositional model for the Earth's core. In *The Mantle and Core*, R. W. Carlson (Ed.). Treatise on Geochemistry, Vol. 2. Elsevier-Pergamon, 547–568.

[5] William F. McDonough and Shih Sueh Sun. 1995. The composition of the Earth. *Chemical Geology* 120, 3–4 (1995), 223–253. doi:10.1016/0009-2541(94)00140-4

[6] Florent Poux and Roland Billen. 2019. Voxel-Based 3D Point Cloud Semantic Segmentation: Unsupervised Geometric and Relationship Featuring vs Deep Learning Methods. *ISPRS International Journal of Geo-Information* 8, 5 (2019), 213. doi:10.3390/ijgi8050213

[7] Abin Shakya, Dipta B. Ghosh, Colin Jackson, Gabriele Morra, and Bijaya B. Karki. 2024. Insights into core-mantle differentiation from bulk Earth melt simulations. *Scientific Reports* 14, 18739 (2024). doi:10.1038/s41598-024-69873-8