# Simultaneous Discovery of Reaction Coordinates and Committor Functions Using Equivariant Graph Neural Networks

**Killian Sheriff**[1,2], **Rodrigo Freitas**[2], **Amalie Trewartha**[1], **Steven Torrisi**[1]
[1]Toyota Research Institute, Los Altos, CA 94022
[2]Department of Materials Science and Engineering,
Massachusetts Institute of Technology, Cambridge, MA 02139
`{steven.torrisi, amalie.trewartha}@tri.global`
`{ksheriff, rodrigof}@mit.edu`

## Abstract

Atoms rearrange themselves during materials synthesis; understanding this self-organization choreography would help the design of novel synthesis recipes. Yet, the mechanisms of such phase transformations are often governed by statistically improbable atomic transitions — known as rare events — that are challenging to investigate by direct, brute-force sampling with conventional atomistic simulations. The transition-state theory framework has been successfully applied for numerous rare-event sampling techniques, which require prior knowledge of reaction coordinates to be encoded in a committor function. Here we show how E(3)-equivariant graph neural networks can be used to simultaneously learn physically appropriate reaction coordinates and committors, solely from molecular dynamics simulations near the start and end states of a reaction. This approach is applied to two dramatically different systems and associated mechanisms, namely the conformational transition in a alanine dipeptide molecule and the solid-liquid transition in the solidification of the CrFeNi metallic alloy. We demonstrated that this approach reduces the need for human intervention in designing reaction coordinates and committor functions, which may enable the high-throughput study of transition states.

## 1 Introduction

Autonomous and high-throughput laboratories promise a reduction in time and labor required for experiments, yet they still lag far behind the rate at which simulations can propose new candidates[1–5]. However, in light of evidence that current thermodynamic heuristics cannot reliably predict successful syntheses[6,7], the community stands to benefit from the development of new atomistic modeling techniques that can guide the synthesis of solid state materials, and enhance synthesis success rates.

In solid-state materials synthesis, it is understood that reactions occur through a sequence of intermediate phases generated from an initial combination of precursors[8,9]. Given that reactions happen at the interfaces of phases (fig. 1a), the number of unique interfaces can grow exponentially, creating a combinatorial challenge[10,11]. The tools of transition state theory allow detailed understanding of reaction dynamics, but are labor intensive and require prior knowledge of the system. Therefore, reducing the labor required to identify and sample rare, rate-limiting interfacial reaction events from molecular dynamics (MD) would be highly valuable to understanding reaction kinetics[12], especially as modern machine-learned interatomic potentials offer long time-and length-scale dynamics with quantum mechanical accuracy[13–20].

38th Conference on Neural Information Processing Systems (NeurIPS 2024 AI4Mat Workshop).
Submitted: September 6th, 2024. Accepted: October 7th, 2024.

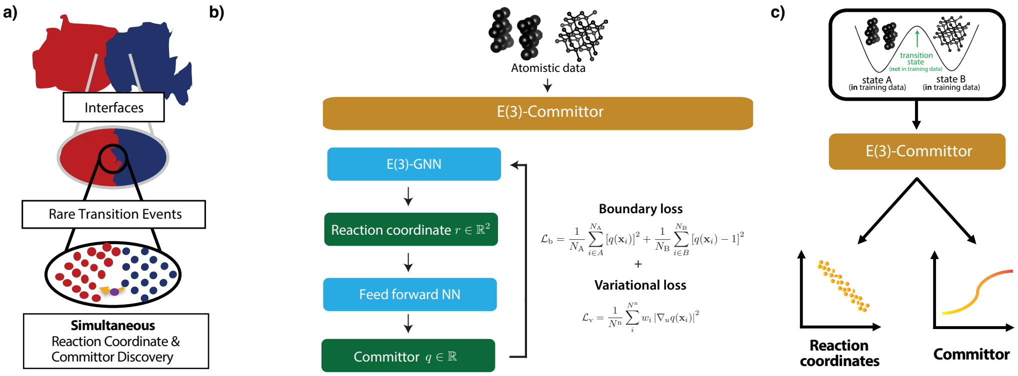

Figure 1: **Simultaneous discovery of reaction coordinates and committor functions. a)** Interfacial dynamics are often dominated by rare transition events. We aim to automate the identification of reaction coordinates and committor functions, minimizing the need for human effort. **b)** Our model architecture processes molecular dynamics data by converting it into a graph representation. We employ an E(3)-equivariant graph neural network (E(3)-GNN) to reduce both structural and chemical information to a lower-dimensional representation, referred to as reaction coordinates. These reaction coordinates are then fed into a feed forward neural network to compute the committor function. The model is optimized to learn both the committor function and reaction coordinates simultaneously through a variational and a boundary loss (see app. A for details). **c)** Our approach identifies physically-appropriate reaction coordinates and committor functions based solely on the dynamics near the initial and final states of the reaction.

Enhanced rare-event sampling has been shown to allow deeper insights into interfacial dynamics[21]. This approach relies on the concept of a *reaction coordinate* ($\mathbf{r}$): a lower-dimensional representation of atomic coordinates that tracks the evolution between two distinct states. A committor function $q(\mathbf{r})$ is employed to map the reaction coordinate to a continuous range between 0 and 1, where values near 0 and 1 respectively correspond to the initial and final states of the system, and a value of 0.5 marks the transition state[22].

Recently, the authors of ref. 23 demonstrated how $q(\mathbf{r})$ could be bootstrapped from atomistic simulations of configurations near the initial and end states of the reaction, and how its gradient magnitude can be used to create biasing potentials that guide the system's time evolution towards the transition state (see app. A for a brief summary of their approach). A key limitation of ref. 23, and of other known approaches for committor function learning[24–26], is the requirement that a reaction coordinate $\mathbf{r}$ be known ahead of time. Identifying physically-appropriate $\mathbf{r}$ requires detailed knowledge of the system's transition, often requiring complex human-designed mathematical functions[21,27,28], rendering rare-event sampling methods unsuitable to high-throughput approaches. Using machine learning for reaction coordinate discovery is becoming more popular[29–31]. The work of Sun *et al.*[32] shows a multi-task framework that identifies low dimensional system descriptors and basin classification that behaves as a committor. They encode system configuration with a multi-layer perceptron-like architecture and use potential energy to help distinguishing reaction progress (which the authors note may not help for diffusion-dominated transitions). Equivariant models may remove both the need for energy to distinguish environments and exploit the efficiency of equivariant networks to learn from atomistic data[14,20,33]. This approach is desirable because it makes no assumptions about either the reaction coordinate or the functional form of the committor.

Here, we enhanced the framework presented in ref. 23 by leveraging an upstream equivariant graph neural network (fig. 1b) to simultaneously identify both reaction coordinates ($\mathbf{r}$) and committor functions ($q$) directly from atomistic simulation data (fig. 1c). We demonstrate the efficacy of our framework to discover physically-appropriate $\mathbf{r}$ and $q(\mathbf{r})$ by applying it to atomistic simulations of the conformational transition of a alanine dipeptide molecule (fig. 2), and of the solid-liquid transitions during solidification of the CrFeNi metallic alloy (fig. 3), and comparing the learned $\mathbf{r}$ and $q(\mathbf{r})$ with known well-established representations.

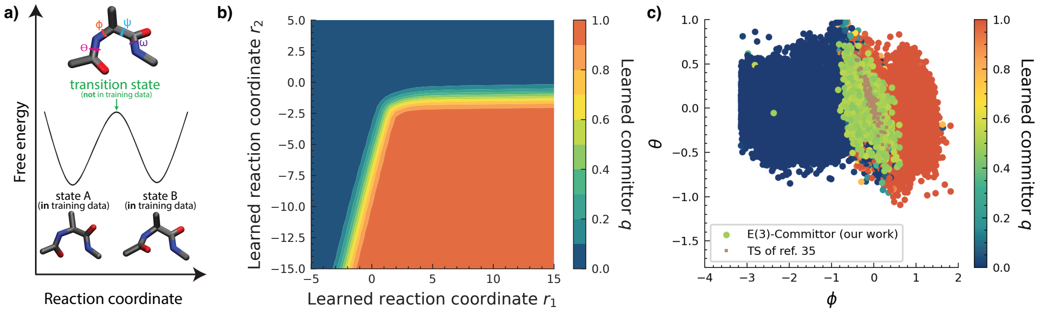

Figure 2: **Learned alanine dipetide committor function.** **a)** Illustration of the conformational space of alanine dipeptide, and of its dihedral angles $\phi, \theta, \psi$, and $\omega$. **b)** Projection of the learned committor function onto the discovered reaction coordinate plane, with committor values of $q(\mathbf{r}_i) \approx 0.5$ identifying the transition state. **c)** Dihedral angles and learned committor value for atomistic simulations undergoing the conformational transition. To enhance visualization, atomistic systems with a learned committor value $q(\mathbf{r}_i) \in [0.45, 0.55]$ were assigned the highest drawing order. Observe the correlation between the transition state (TS) dihedral angles, digitized from ref. 35, and the dihedral angles of atomistic systems where our method assigns a committor value of $q(\mathbf{r}_i) \approx 0.5$. Our framework accurately identifies a physically relevant transition state ($\theta \approx -\phi$) directly from the structure, without prior knowledge of the transition states, and of dihedral angles.

## 2 Results

### 2.1 Learned alanine dipeptide committor function

Alanine dipeptide is a widely used benchmark molecule for studying transition states[23,32,34]. Its molecular conformation states (fig. 2a) have well-defined energy minima[28] described by the dihedral angles $\phi, \theta, \psi$, and $\omega$[23]. Notably, configurations at the transition state between conformations $C_{7eq}$ (state A) and $C_{7ax}$ (state B)[1] are characterized by the inverse relationship $\theta \approx -\phi$[23,28,35].

Here, we showcase how our model discovers a committor function and corresponding reaction coordinate of alanine dipetide, by training exclusively on MD data from states A and B without any prior knowledge of the transition state or expert-derived reaction coordinate. The first step in our process involves converting the unbiased atomistic data of states A and B (collected in ref. 23) into a graph representation, which is subsequently inserted into our committor network (see sec. 4.1 for details). Our framework enables simultaneous learning of both the reaction coordinates $\mathbf{r}$ and the committor function $q(\mathbf{r})$. Figure 2b shows the discovered reaction coordinate space; notice that a clear transition state is identified with a committor value $q(\mathbf{r}_i) \approx 0.5$. Next, we investigate whether this transition state corresponds to the correct physical transition state of alanine dipeptide. This is accomplished by comparing in fig.2c the dihedral angles and learned committor of atomistic simulations undergoing the conformational change (taken from the biased atomistic simulations of ref. 23), to the dihedral angles from ref. 35, which were obtained using well-established rare-event sampling techniques. While it is clear that there is a strong correlation between our discovered committor and previously-known reaction coordinates, a more definitive validation would involve running atomistic simulations with a biasing potential derived directly from our learned committor — an approach we plan to explore in future work. Nevertheless, our findings suggest that it is possible to learn relevant committors and reaction coordinates directly from the atomistic data near the start and end states.

### 2.2 Learned CrFeNi solid-liquid committor function

We next turn to the study of materials synthesis by applying our method to some 100,000 atom MD simulations of solidification from a spherical crystalline seed of the CrFeNi metallic alloy (fig. 3a, see sec. 4.2 for details). The kinetics of the atomic transitions from the all-liquid (state A) to the all-solid phase (state B) is governed by rare events[36,37]. Figure 3b compares the learned committor with the

---

[1] So named due to the orientation of the $C_\beta$ atom being in an axial or equatorial orientation relative to a 7-atom ring formed by intramolecular hydrogen bonding[28].

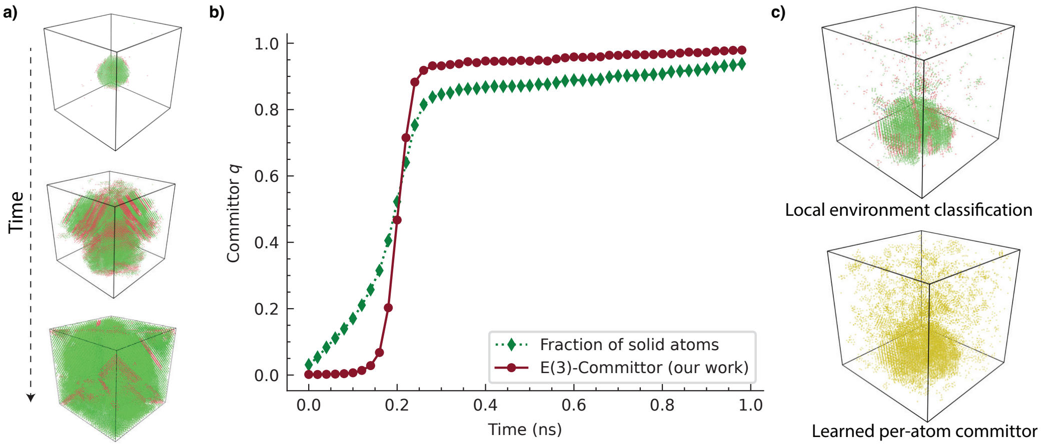

Figure 3: **Learned CrFeNi committor function. a)** Crystal growth from a spherical CrFeNi seed, with atoms colored by their local crystal structure (see app. B for details). **b)** The learned committor over simulation time approximates the fraction of solid atoms in the system. **c)** Comparison of local atomic environment classification (top) and per-atom committor values (bottom) after removing the network's pooling layers post-training. Atoms identified as liquid (top) and with $q \leq 0.9999$ (bottom) are hidden.

fraction of solid atoms in the simulation, which was obtained from a crystal-structure classifier known as polyhedral template matching[38] (see app. B for details). Notice the close proximity between the curves, showing that from structure, we were able to derive a physically relevant global description of the system state which captures the increase of solid-like atomic environments. Addtionally, the correlation of pre-pooling values with local descriptors of crystallinity (fig. 3c) suggests that the learned committor gradients, propagated backwards through the reaction coordinates, could direct atomic movements towards solid or liquid-like committor-derived biasing potentials.

Note that unlike in the molecular case, where the conformation is a global variable, it appears that the global system state is roughly tracking the fraction of atoms that are in the solid versus liquid phase. This, and the fact that our formalism requires the identification of only two distinct states, raises the questions of how to proceed in (1) precisely identifying the transition state for local phase transitions, and (2) how to study complex multi-step processes that involve intermediate species or by-products. We leave deeper study of both of these for future work, but propose here some initials thoughts. For (1), two possible ways to do this are to study the contribution of individual atoms to the committor along the natural trajectory, or to use a biasing potential to examine the trajectory that maximizes an atom's contribution to the committor. We hope that this procedure may reduce the labor of identifying and studying a transition state given molecular dynamics data (e.g., do atoms with intermediate local contributions to the committor represent those undergoing local transformations? Could these local transformations provide insights into the mechanisms driving variations in growth behavior?). For (2), multi-step processes, we do not know a priori how intermediate or byproduct phases could interact with this framework. However, making it easier to generate intermediate steps could enable easier sampling campaigns while zooming in to proposed transition states, iteratively applying this framework to successively smaller in-between steps in a more complex phase transition. However, further study would need to be conducted to estimate the risk of a non rate-limiting transition pathway to be picked up by the model.

Optimizing computational performance is crucial for realistic simulations of materials processing and synthesis, which often require models capable of handling millions of atoms per time step[16]. Enhancing the inference speed of our E(3)-GNN (currently operating at approximately 14 seconds per hundred thousand atom without just-in-time compilation) should not be overlooked. Integrating eSCN layers[39] offers an effective solution to reduce computational complexity by replacing $SO(3)$ convolutions with mathematically equivalent $SO(2)$ convolutions. Additionally, investigating alternative equivariance methods, such as EGNN's use of relative atomic distances[40], could provide insights into the differences in representational richness and computational complexity. Finally, examining the role of symmetry-inductive bias could highlight its benefits in low-data regimes[14,41] and determine if the learned coordinates and committor display greater variability without it, thereby justifying the additional computational complexity.

Consequently, future work will extend to solid-solid phase transitions[21], examine how variations in loss functions and system states affect the learned reaction coordinates (see app. C for preliminary intuitions), and will benchmark our framework against other methods by comparing transition state sampling performances.

## 3 Conclusion

In this work we show evidence that E(3)-equivariant neural networks can be used to simultaneously identify reaction coordinates and committor functions solely from atomistic simulation data at the start and end states of a transition of interest. We anticipate that this approach may simplify and help automate the discovery of physically-appropriate reaction coordinates and committor functions, which can facilitate the identification and understanding of rate-limiting transition states associated with interfacial reactions common during solid-state synthesis[8]. Facilitating the rapid and automatic modeling of interfacial dynamics may accelerate the refinement of experimental parameters and enhance synthesis success rates.

## 4 Methods

### 4.1 E(3)-equivariant committor network

Molecular dynamics snapshots at every time step $t$ are transformed into a graph representation, which is then processed by an $\mathbb{E}(3)$-equivariant graph neural network to produce a lower-dimensional representation of the system state $\mathbf{r}^t \in \mathbb{R}^2$. Our architecture, inspired from refs. 20, 33, 42, is implemented using the e3nn package[43]. The network consists of two $\mathbb{E}(3)$-equivariant convolutions, which employ $\mathcal{O}(3)$-equivariant filters constructed using spherical harmonics $Y_\ell^m(r)$ up to degree $\ell_{\text{max}} = 1$, and of a radial basis network, composed of ten cosine radial basis functions. The hidden layer features transform according to the irreducible representation $25\times 1o + 25\times 1e + 25\times 0o + 25\times 0e$, where "$\ell p$" denotes the e3nn data types, with $\ell$ being the angular frequency, and $p$ representing parity ($p = o$ for odd and $p = e$ for even). The lower dimensional representation of the system state $z_t \in \mathbb{R}^2$ is then fed to a feed-forward neural network composed of three linear layers with output dimension 32, 32, and one respectively. Hyperbolic tangent functions were used as activation function in the feed-forward neural network, given that the loss function (see app. A for details) depends on the derivative of the inputs. Additionally, to facilitate the learning of the committor, a sharp sigmoid-like activation function is used in the last layer.

### 4.2 Solidification simulation

Solidification simulations of the equiatomic CrFeNi medium entropy alloy were carried out using the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) software[44], with the EAM potential of ref. 45. The initial structure was composed of 100,000 atoms, with 2,400 atoms arranged as a spherical crystalline seed in a face-centered cubic structure, while the remaining atoms were distributed randomly to mimic the liquid phase. The system was equilibrated in two steps. First, the liquid atoms were relaxed using a conjugate gradient algorithm for 100 steps. Then, the liquid atoms were equilibrated for 1 ps at zero pressure and finite temperature, using a Nosé-Hoover thermostat and barostat with a 0.1 ps temperature damping parameter, a 1 ps pressure damping parameter, and a chain length of three, allowing isotropic expansion and contraction. The crystalline seed atoms remained fixed in their initial structure with constant lattice parameters, unaffected by the liquid atoms' behavior. The growth is then carried on using a Nosé-Hoover thermostat and barostat with a 1 ps damping parameter and a chain length of 3 to maintain the system at finite temperature (800 K) and zero pressure for up to 1 ns.

## References

[1] Chi Chen and Shyue Ping Ong, "A universal graph deep learning interatomic potential for the periodic table", *Nature Computational Science* (Nov. 2022), ISSN: 2662-8457, DOI: `10.1038/s43588-022-00349-3`.

[2] Joseph H. Montoya, Kirsten T. Winther, Raul A. Flores, Thomas Bligaard, Jens S. Hummelshøj, and Muratahan Aykol, "Autonomous intelligent agents for accelerated materials discovery", *Chem. Sci.* (32 2020), DOI: `10.1039/D0SC01101K`.

[3] Amil Merchant, Simon Batzner, Samuel S. Schoenholz, Muratahan Aykol, Gowoon Cheon, and Ekin Dogus Cubuk, "Scaling deep learning for materials discovery", *Nature* (Nov. 2023), ISSN: 1476-4687, DOI: `10.1038/s41586-023-06735-9`.

[4] Akshay Subramanian, Wenhao Gao, Regina Barzilay, Jeffrey C. Grossman, Tommi Jaakkola, Stefanie Jegelka, Mingda Li, Ju Li, Wojciech Matusik, Elsa Olivetti, Connor W. Coley, and Rafael Gomez-Bombarelli, "Closing the Execution Gap in Generative AI for Chemicals and Materials: Freeways or Safeguards", *An MIT Exploration of Generative AI* (Mar. 2024), https://mit-genai.pubpub.org/pub/681kpeoa.

[5] Elton Pan, Christopher Karpovich, and Elsa Olivetti, *Deep Reinforcement Learning for Inverse Inorganic Materials Design*, 2022, arXiv: `2210.11931 cond-mat.mtrl-sci`.

[6] Awadhesh Narayan, Ankita Bhutani, Samantha Rubeck, James N Eckstein, Daniel P Shoemaker, and Lucas K Wagner, "Computational and experimental investigation for new transition metal selenides and sulfides: the importance of experimental verification for stability", *Physical Review B* (2016).

[7] Joseph H Montoya, Carolyn Grimley, Muratahan Aykol, Colin Ophus, Hadas Sternlicht, Benjamin H Savitzky, Andrew M Minor, Steven B Torrisi, Jackson Goedjen, Ching-Chang Chung, et al., "How the AI-assisted discovery and synthesis of a ternary oxide highlights capability gaps in materials science", *Chemical Science* (2024).

[8] Muratahan Aykol, Joseph H Montoya, and Jens Hummelshøj, "Rational solid-state synthesis routes for inorganic materials", *Journal of the American Chemical Society* (2021).

[9] Jiadong Chen, Samuel R Cross, Lincoln J Miara, Jeong-Ju Cho, Yan Wang, and Wenhao Sun, "Navigating phase diagram complexity to guide robotic inorganic materials synthesis", *Nature Synthesis* (2024).

[10] Akira Miura, Christopher J Bartel, Yosuke Goto, Yoshikazu Mizuguchi, Chikako Moriyoshi, Yoshihiro Kuroiwa, Yongming Wang, Toshie Yaguchi, Manabu Shirai, Masanori Nagao, et al., "Observing and modeling the sequential pairwise reactions that drive solid-state ceramic synthesis", *Advanced Materials* (2021).

[11] Matthew J McDermott, Brennan C McBride, Corlyn E Regier, Gia Thinh Tran, Yu Chen, Adam A Corrao, Max C Gallant, Gabrielle E Kamm, Christopher J Bartel, Karena W Chapman, et al., "Assessing thermodynamic selectivity of solid-state reactions for the predictive synthesis of inorganic materials", *ACS Central Science* (2023).

[12] Mingjian Wen, Evan Walter Clark Spotte-Smith, Samuel M Blau, Matthew J McDermott, Aditi S Krishnapriyan, and Kristin A Persson, "Chemical reaction networks and opportunities for machine learning", *Nature Computational Science* (2023).

[13] Jonathan Vandermause, Steven B Torrisi, Simon Batzner, Yu Xie, Lixin Sun, Alexie M Kolpak, and Boris Kozinsky, "On-the-fly active learning of interpretable Bayesian force fields for atomistic rare events", *npj Computational Materials* (2020).

[14] Simon Batzner, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P Mailoa, Mordechai Kornbluth, Nicola Molinari, Tess E Smidt, and Boris Kozinsky, "E (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials", *Nature communications* (2022).

[15] Ilyes Batatia, Simon Batzner, Dávid Péter Kovács, Albert Musaelian, Gregor NC Simm, Ralf Drautz, Christoph Ortner, Boris Kozinsky, and Gábor Csányi, "The design space of e (3)-equivariant atom-centered interatomic potentials", *arXiv preprint arXiv:2205.06643* (2022).

[16] Mahmudul Islam, Killian Sheriff, Yifan Cao, and Rodrigo Freitas, "Nonequilibrium chemical short-range order in metallic alloys" (2024), arXiv: `2409.15474 cond-mat.mtrl-sci`.

[17] Ilyes Batatia, David P Kovacs, Gregor Simm, Christoph Ortner, and Gábor Csányi, "MACE: Higher order equivariant message passing neural networks for fast and accurate force fields", *Advances in Neural Information Processing Systems* (2022).

[18] Yifan Cao, Killian Sheriff, and Rodrigo Freitas, *Capturing short-range order in high-entropy alloys with machine learning potentials*, 2024, arXiv: `2401.06622 cond-mat.mtrl-sci`.

[19] Anders Johansson, Yu Xie, Cameron J Owen, Jin Soo Lim, Lixin Sun, Jonathan Vandermause, and Boris Kozinsky, "Micron-scale heterogeneous catalysis with Bayesian force fields from first principles and active learning", *arXiv preprint arXiv:2204.12573* (2022).

[20] Killian Sheriff, Yifan Cao, Tess Smidt, and Rodrigo Freitas, "Quantifying chemical short-range order in metallic alloys", *Proceedings of the National Academy of Sciences* (2024), DOI: `10.1073/pnas.2322962121`.

[21] Yanyan Liang, Grisell Díaz Leines, Ralf Drautz, and Jutta Rogal, "Structural transformations driven by local disorder at interfaces", *Physical Review Materials* (2024).

[22] A Kolmogorov, "Über die analytischen Methoden in der Wahrscheinlichkeitstheorie", *Math Annal* (1931).

[23] Peilin Kang, Enrico Trizio, and Michele Parrinello, "Computing the committor with the committor to study the transition state ensemble", *Nature Computational Science* (2024).

[24] Qianxiao Li, Bo Lin, and Weiqing Ren, “Computing committor functions for the study of rare events using deep learning”, *The Journal of Chemical Physics* (2019).

[25] Yi Isaac Yang, Qiang Shao, Jun Zhang, Lijiang Yang, and Yi Qin Gao, “Enhanced sampling in molecular dynamics”, *The Journal of Chemical Physics* (Aug. 2019), ISSN: 0021-9606, DOI: `10.1063/1.5109531`, eprint: `https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/1.5109531/19987375/070902\_1\_1.5109531.pdf`.

[26] Jérôme Hénin, Tony Lelièvre, Michael R. Shirts, Omar Valsson, and Lucie Delemotte, “Enhanced Sampling Methods for Molecular Dynamics Simulations [Article v1.0]”, *Living Journal of Computational Molecular Science* (Dec. 2022), DOI: `10.33011/livecoms.4.1.1583`.

[27] Maya M Martirossyan, Matthew Spellings, Hillary Pan, and Julia Dshemuchadse, “Local structural features elucidate crystallization of complex structures”, *ACS nano* (2024).

[28] Vladimir Mironov, Yuri Alexeev, Vikram Khipple Mulligan, and Dmitri G Fedorov, “A systematic study of minima in alanine dipeptide”, *Journal of Computational Chemistry* (2019).

[29] Baron Peters, “Using the histogram test to quantify reaction coordinate error”, *The Journal of chemical physics* (2006).

[30] Hendrik Jung, Roberto Covino, and Gerhard Hummer, “Artificial intelligence assists discovery of reaction coordinates and mechanisms from molecular dynamics simulations”, *arXiv preprint arXiv:1901.04595* (2019).

[31] Luigi Bonati, Enrico Trizio, Andrea Rizzi, and Michele Parrinello, “A unified framework for machine learning collective variables for enhanced sampling simulations: mlcolvar”, *The Journal of Chemical Physics* (2023).

[32] Lixin Sun, Jonathan Vandermause, Simon Batzner, Yu Xie, David Clark, Wei Chen, and Boris Kozinsky, “Multitask machine learning of collective variables for enhanced sampling of rare events”, *Journal of Chemical Theory and Computation* (2022).

[33] Killian Sheriff, Yifan Cao, and Rodrigo Freitas, “Chemical-motif characterization of short-range order with E(3)-equivariant graph neural networks”, *npj Computational Materials* (Sept. 2024), ISSN: 2057-3960, DOI: `10.1038/s41524-024-01393-5`.

[34] Jiri Vymetal and Jiri Vondrasek, “Metadynamics As a Tool for Mapping the Conformational and Free-Energy Space of Peptides - The Alanine Dipeptide Case Study”, *The Journal of Physical Chemistry B* (2010).

[35] Peter G. Bolhuis, Christoph Dellago, and David Chandler, “Reaction coordinates of biomolecular isomerization”, *Proceedings of the National Academy of Sciences* (2000), DOI: `10.1073/pnas.100127697`, eprint: `https://www.pnas.org/doi/pdf/10.1073/pnas.100127697`.

[36] Emily B Moore and Valeria Molinero, “Structural transformation in supercooled water controls the crystallization rate of ice”, *Nature* (2011).

[37] Rodrigo Freitas and Evan J Reed, “Uncovering the effects of interface-induced ordering of liquid on crystal growth using machine learning”, *Nature communications* (2020).

[38] Peter Mahler Larsen, Søren Schmidt, and Jakob Schiøtz, “Robust structural identification via polyhedral template matching”, *Modelling and Simulation in Materials Science and Engineering* (2016).

[39] Saro Passaro and C. Lawrence Zitnick, *Reducing SO(3) Convolutions to SO(2) for Efficient Equivariant GNNs*, 2023, arXiv: `2302.03655 cs.LG`.

[40] Victor Garcia Satorras, Emiel Hoogeboom, and Max Welling, “E(n) Equivariant Graph Neural Networks”, *CoRR* (2021), arXiv: `2102.09844`.

[41] Emiel Hoogeboom, Victor Garcia Satorras, Clément Vignac, and Max Welling, *Equivariant Diffusion for Molecule Generation in 3D*, 2022, arXiv: `2203.17003 cs.LG`.

[42] Joshua A. Rackers, Lucas Tecot, Mario Geiger, and Tess E Smidt, “A recipe for cracking the quantum scaling limit with machine learned electron densities”, *Machine Learning: Science and Technology* (2023), DOI: `10.1088/2632-2153/acb314`.

[43] Mario Geiger and Tess Smidt, *e3nn: Euclidean Neural Networks*, 2022, arXiv: `2207.09453 cs.LG`.

[44] A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. in ’t Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, R. Shan, M. J. Stevens, J. Tranchida, C. Trott, and S. J. Plimpton, “LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales”, *Comp. Phys. Comm.* (2022), DOI: `10.1016/j.cpc.2021.108171`.

[45] Laurent Karim Béland, Artur Tamm, Sai Mu, German D. Samolyuk, Yuri N. Osetsky, Alvo Aabloo, Mattias Klintenberg, Alfredo Caro, and Roger E. Stoller, “Accurate classical short-range forces for the study of collision cascades in Fe–Ni–Cr”, *Computer Physics Communications* (2017), ISSN: 0010-4655, DOI: `https://doi.org/10.1016/j.cpc.2017.05.001`.

[46] Zhouyang Zhang, Yujie Tang, Yiran Ying, Junqing Guo, Min Gan, Yateng Jiang, Chunxian Xing, Shanshan Pan, Ming Xu, Yangbo Zhou, et al., “Multistep nucleation visualized during solid-state crystallization”, *Materials Horizons* (2022).

[47] Baron Peters, “Reaction coordinates and mechanistic hypothesis tests”, *Annual review of physical chemistry* (2016).

## A Committor loss function

To briefly summarize the approach of ref. 23, they propose learning the committor function by simultaneously optimizing two types of losses: the boundary condition loss $\mathcal{L}_b$ (eq. 1) and the variational loss $\mathcal{L}_v$ (eq. 2). The boundary condition loss $\mathcal{L}_b$ ensures that the reaction coordinates of the initial state “A” and final state “B” are mapped to 0 and 1 respectively. The variational loss $\mathcal{L}_v$ acts as a regularizer, smoothing the transition between the initial and final states in the reaction coordinate space.

Mathematically, those are defined as:

$$\mathcal{L}_\mathrm{b} = \frac{1}{N_\mathrm{A}}\sum_{i\in A}^{N_\mathrm{A}}[q(\mathbf{r}_i)]^2 + \frac{1}{N_\mathrm{B}}\sum_{i\in B}^{N_\mathrm{B}}[q(\mathbf{r}_i)-1]^2 \tag{1}$$

$$\mathcal{L}_\mathrm{v} = \frac{1}{N^\mathrm{n}}\sum_{i}^{N^\mathrm{n}} w_i \left|\nabla_u q(\mathbf{r}_i)\right|^2 , \tag{2}$$

where $q$ is the committor function of a system characterized by its reaction coordinate $\mathbf{r}_i$, $N_\mathrm{A}$,$N_\mathrm{B}$ are the number of atomistic systems in states A and B, respectively. $N^\mathrm{n}$ is the number of systems initiated at iteration n, and $w_i$ is a weighting factor which accounts for the gradual sampling of the transition state ensemble as the iterative process of running atomistic simulations with an increasingly refined biasing potential progresses.

The committor loss function $\mathcal{L}$ is expressed as:

$$\mathcal{L} = \mathcal{L}_\mathrm{v} + \alpha\mathcal{L}_\mathrm{b},$$

where $\alpha$ modulates the relative importance of the two loss.

Finally, a biasing potential $V_b$ can be derived from the learned committor

$$V_\mathrm{b} = -\frac{1}{\beta} log\left(\left|\nabla_u q(\mathbf{r}_i)\right|^2\right),$$

where $\beta$ is the inverse temperature. Based on the overall behavior of $q(x)$, $V_\mathrm{b}$ is repulsive in bassin A and B, given that $\nabla q(x) \approx 0$ there. However, near the transition state, where $q(x)$ increases sharply from 0 to 1, $V_\mathrm{b}$ becomes significantly attractive.

We note that this procedure assumes two distinct states A and B, and our paper exclusively assumes two-step processes. In real synthesis processes, sometimes reactions occur via multi-step procedures, but we speculate that the enumeration and study of possible intermediate states could be an application of this method[46].

## B Local crystal structure identification

Local crystal structures were identified using the polyhedral template matching method[38] with a cutoff of 0.08. Red indicates hexagonal close-packed environments, green indicates face-centered cubic environments, and liquid environments are omitted for clarity.

## C Influence of the variational loss onto the reaction coordinate space

We hope to show that the benefit of the variational term in the committor loss (eq. 2) is to help the network learn reaction coordinates suited to lower-dimensional mapping, effectively regularizing the learned coordinates. However, we emphasize that these observations remain speculative as we are still in the process of validating this hypothesis. The precise interpretation of the learned descriptions are still under study, as the crystal growth has already begun before the value attains $q(\mathbf{r}_i) \approx 0.5$, suggesting that the learned committor is describing mostly the population-level statistics of atoms; the formulation of the committor as describing the probability of a configuration transitioning to state B[47] may not readily apply here.

## D Code and data availability

The code and data used for this project is available at https://github.com/TRI-AMDD/interstate.