ML Force Fields for Computational NMR Spectra of Dynamic Materials across Time-Scales

Lars L. Schaaf *
University of Cambridge

Benjamin J. Rhodes *
University of Cambridge

Mary E. Zick
Cornell University

Suzi M. Pugh
University of Cambridge

Jordon S. Hilliard
The Ohio State University

Shivani Sharma
University of Cambridge

Casey R. Wade
The Ohio State University

Phillip J. Milner
Cornell University

Gábor Csányi
University of Cambridge

Alexander C. Forse
University of Cambridge
`acf50@cam.ac.uk`

Abstract

Automated material discovery relies on the ability to accurately characterize synthesized materials, a task where solid-state nuclear magnetic resonance (NMR) spectroscopy plays a crucial role due to its atom-level insights. However, interpreting solid-state NMR spectra is challenging, often requiring quantum mechanical calculations that traditionally assume static materials. This assumption neglects the dynamic nature of materials at operational temperatures, leading to discrepancies between computational and experimental results. To overcome this limitation, we propose an approach that incorporates both molecular dynamics and transition state searching to model dynamic effects across timescales in NMR spectra using machine learning force fields (MLFFs). By fine-tuning atomistic foundation models, we achieve accurate and cost-effective MLFFs in an automated manner. We validate our approach through ^{17}O NMR experiments on porous materials for carbon capture. Specifically we look at the metal-organic framework MFU-4l, demonstrating that existing methods for predicting spectra fail to match experimental observations, while our method achieves strong agreement. This workflow not only facilitates automated characterization of materials critical for carbon capture but also highlights the potential of ML-driven simulations in predicting material properties.

1 Introduction

Accurately differentiating materials and determining their atomic structures is crucial for automated characterizations in autonomous laboratories. Nuclear magnetic resonance (NMR) spectroscopy provides high-resolution insights into a material's atomic structure, making it essential for understanding processes from carbon capture to catalysis. However, solid-state NMR presents unique challenges due to its complexity in interpretation. Expensive, quantum mechanical methods, such as density functional theory (DFT) calculations, are often required to match the experimental result with a materials atomic arangement. A single DFT NMR prediction represents a spectrum of a frozen configuration, neglecting all dynamical effects.

38th Conference on Neural Information Processing Systems 2024. Accelerated Materials Design Workshop.

Machine learning force fields (MLFFs) have been applied to simulate dynamic effects in NMR spectra [1–3]. In these approaches, configurations from molecular dynamics simulations with a MLFF are evaluated with DFT NMR calculations and averaged to obtain effective spectra. Although MLFFs provide computational speedup, they typically capture only fast dynamics (nanosecond timescales). However, longer time-scale dynamics (e.g. microseconds) can also influence NMR shifts, limiting the effectiveness of traditional approaches. Alternatively, packages allow modeling of stochastic hops between a small number of configurations and associated chemical shifts at arbitrary rates [4]. As the number of relevant minima increases the number of required barrier calculations grows combinatorially. Furthermore, the procedure doesn't correspond to a true Boltzmann sampling of relevant equilibrium configuration and hence often fail to capture complex short-term processes accurately. Furthermore, obtaining rates for these slow processes requires expensive ab-initio transition state searching and vibrational analysis.

Our work presents a combined approach that integrates fast- and long-timescale dynamics, enabling agreement between computational and experimental spectra. Furthermore, we present a comprehensive workflow on how to fine-tune atomistic foundation models and use them to both sample equilibrium configurations at short time scales (ns) as-well as calculate transition rates for more rare events. We show that it is possible to obtain accurate computational NMR spectra even for materials with complex dynamics. For a full-text version of the paper targeted at a Chemsitry audience please see Reference [5].

2 Background

NMR spectroscopy The atomic structure of a material can be probed by analyzing how atomic nuclei interact with an external magnetic field. As this interaction strongly depends on the electronic environment surrounding the nucleus, these spectra are a type of fingerprint for a given material.

Ab-initio NMR spectra To relate NMR spectra to atomic environments it is often necessary to perform quantum mechanical calculations. These compute the electron density, under some approximations such as density functional theory (DFT), and its effect on the magnetic shielding tensors of the surrounding nuclei [6–9].

ML Force fields Recent advances in geometric deep learning have led to the development of machine learning force fields (MLFFs), which can predict atomic forces and energies based on 3D atomic configurations with near-quantum mechanical accuracy [10–17]. These can be used as surrogate models to sample relevant configurations using molecular dynamic simulations.

3 Results

In this section, we outline the proposed automated workflow for obtaining dynamic NMR spectra across time scales. Furthermore, we test the approach on novel experimental ^{17}O-NMR-spectra of a carbon capture material, namely the MFU-4l metal organic framework.

3.1 Automated training of ML forcefield

Recently published atomistic foundation models, such as MACE-MP0 [18], can be used to obtain qualitatively correct simulations of a wide variety of material systems. However, to accurately determine the rates of rare events, we find it necessary to fine-tune [19–21]. We have automated this workflow for quickly creating a small data set on the order of 100 configurations that is sufficient to accurately model complex systems like metal organic frameworks. The workflow, outlined in Figure 2, entails using the foundation model to run simulations, such as geometry optimization, global minima searches, or transition state searches. As in Reference [22], we use an active learning workflow [23–25] to enhance our training data in an efficient manner. configurations are sampled from the simulations and reevaluated using an accurate ground truth quantum mechanical calculation. The resulting configurations are added to a training set and the foundation model is fine-tuned. This loop continues until the model reaches the desired accuracy.

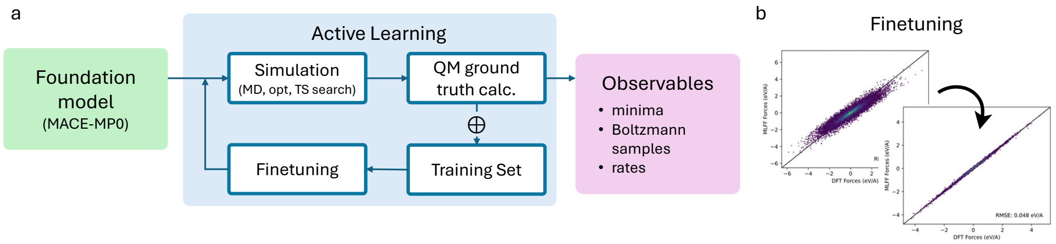

Figure 1: **Automated active learning workflow for fine-tuning foundation models.** Showing how the new configurations for the training set are sampled from atomistic simulations to interactively extend the dataset (a). Figure b shows the correlation between ground truth forces against the foundations models (left) and fine-tuned modeled (right) forces, for respectively unseen configurations.

3.2 Handsfree exploration of dynamic events

Firstly, we explore the potential energy landscape of our material using a mixture of molecular dynamics and global optimization methods. Specifically, we perform minima hopping [26], which consists of alternating NVE molecular dynamics with geometry optimization to find relevant minima. This makes it possible to explore the configuration space beyond the limits of direct molecular dynamics or other thermodynamic methods. Transition-state searches are used to compute rates between minima, enabling us to model rare dynamic events.

3.3 Computing dynamic NMR spectra

Because NMR spectra are sensitive to chagnes in chemical environments of nuclei, they are strongly affected my molecular motion, often referred to as chemical exchange. Indeed, the effect of dynamic processes on the observed NMR spectra depends significantly on the relationship of the process' exchange rate, k, and the absolute resonant frequency difference, $\Delta\nu$, between the initial and final states. This separates, the effect of chemical exchange into three regimes: slow exchange (k $\ll \Delta\nu$), where each environment is well resolved in the NMR spectrum; fast exchange (k $\gg \Delta\nu$), where rapid dynamics lead to a complete averaging of individual NMR environments; or intermediate exchange (k $\approx \Delta\nu$) when NMR environment coalescence is observed but averaging is only partial. See Figure 2, for examples of such coalescence (right).

Using the molecular dynamics samples and transition states, obtained as described in the previous section, we can compute NMR spectra across chemical excahnge regimes, as visually outlined in 2 to the left.

Fast regime We use direct MLFF molecular dynamics to capture fast processes. The final spectrum, Figure 2**C**, is obtained by point-wise averaging of the individual DFT shielding tensors across samples of nanosecond-long trajectories.

Intermediate regime To capture longer timescale events we do an extensive local minima energy structure search using minima hopping [26]. We find the exchange rates connecting the minima using a transition-state search with the MLFF. We use the stochastic hopping package EXPRESS,[4] to compute the effective NMR shielding tensors from the initial and final minimum's DFT shielding tensor and MLFF computed exchange rate.

Fast + Intermediate regime To combine the effects of fast and intermediate exchange processes, we first sample configurations in two minima and compute the two MD-averaged tensors as in Figure 2**C**. We then use EXPRESS to model a stochastic hopping between MD averaged tensors as illustrated in Figure 2**D**. that can be transferred to other similar metal-binding systems.[27]

MFU-4l MOF for carbon capture We now apply this approach to understand CO_2 adsorption in a promising metal organic framework (MOF) as illustrated in Figure 1. Here, the material is dosed with special CO_2, where the oxygen's are enriched ^{17}O to be detectable by NMR. As the natural abundance of the ^{17}O isotope is low (less than 0.04%), the peaks correspond almost exclusively to oxygen atoms originating from enriched CO_2 [28]. This gives us a unique way to track what happens to CO_2 as it is adsorbed by the MOF. When the MFU-4l MOF is dosed with carbon dioxide, it

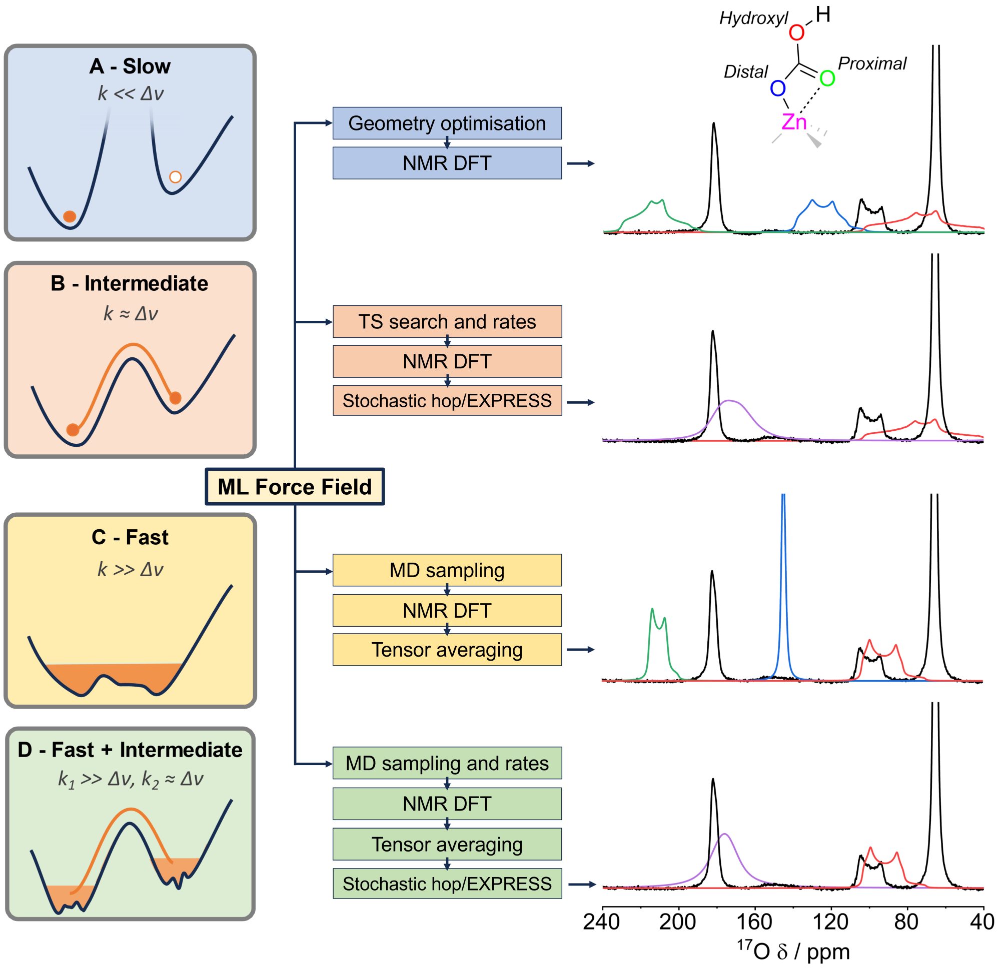

Figure 2: **Capturing dynamic effects of NMR spectra across timescales** The ground truth experimental ^{17}O NMR spectrum (right, black) of MFU-4l MOF dosed with carbon dioxide. Computational NMR spectra associated with the individual oxygen atoms are shown in color. Schematics of the exchange regimes and workflows utilising ML force-fields are displayed (left). Panel D demonstrates the most accurate model combining both fast and intermediate exchange modelling.

reacts with the MOF forming bicarbonate. This is immediately visible in the experimental spectrum, to the right of Figure 2 in black. The experimental spectrum has three peaks; the far right peak corresponds to carbon dioxide, which has not reacted with the MOF and is physisorbed. We focus only on modeling the two interesting peaks, which correspond to CO_2 strongly chemically adsorbed by the MOF. As visible from the **static picture (A)**, there are 3 distinct chemisorbed peaks in the computed NMR spectrum (in color), corresponding to the three oxygen atoms of the bicarbonate. The **intermediate regime (B)** reveals why the experimental spectrum only has two chemisorbed peaks. We find that the bicarbonate proton can rotate around its bonded oxygen atom, effectively causing the two oxygen environments to be identical, as illustrated in Figure 3. A transition-state search finds the barrier to be 0.49 eV. Combined with a vibrational analysis of the start and transition state configuration, the predicted reaction rate corresponds to 2.4 x 10^5 Hz under the harmonic approximation at room temperature. The EXPRESS[4] calculation shows that this rate causes a coalescence of the two peaks. However, the computational peaks are too wide, especially the peak at 90 ppm. Secondly, a nanosecond-long MD simulation reveals a rotational tumbling of the bounded bicarobnate. Including the effects of these **fast regime (C)** dynamics leads to a reduction in the average anisotropic chemical shielding resulting in narrower line widths, particularly visible at 90 ppm. Only by combining the **fast and intermediate approach (D)**, we obtain strong agreement with the two chemi-sorbed experimental peaks.

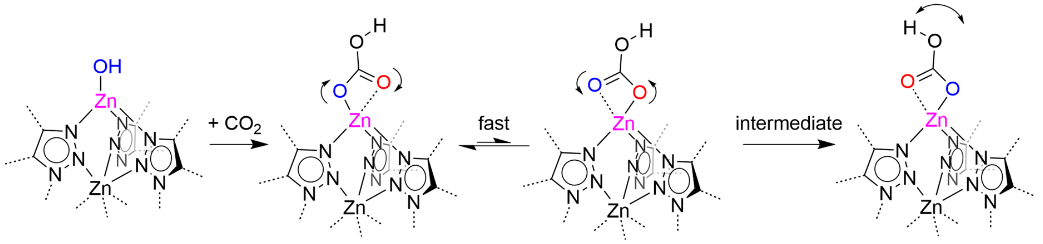

Figure 3: **Dynamics of adsorbed CO_2 in MFU-4l MOF accross chemical exchange regimes.** A cartoon representation of the adsorption process from carbon-dioxide to bicarbonate, rotational tumbling in the fast regime and the rotation of a hydroxyl group in the intermediate regime.

4 Discussions

The automated characterization of complex materials requires experimental techniques that capture atomic-level detail. For many systems, the use of nuclear magnetic resonance (NMR) spectroscopy may be limited due to the difficulty in interpreting experimental spectra, primarily due to the dynamic nature of materials. Our approach automates the incorporation of dynamic effects into computational NMR spectra, utilizing cost-effective fine-tuning of expressive atomistic foundation models to accurately capture dynamics across multiple timescales. By combining tensor averaging with stochastic hopping methods, we achieve strong agreement between computational and experimental results. Although demonstrated on a metal-organic framework for carbon capture, this method offers a generalizable solution for dynamic materials. The workflow opens new opportunities for automated characterization, enabling large-scale material screening and advancing automated material design. To reduce the computational cost of this approach, steps could be taken to learn the magnetic tensorial properties [29], potentially as a future foundation model.

References

1. Fykouras, K. *et al.* Disorder to order: how halide mixing in $\mathrm{MAPbI}_{3-x}\mathrm{Br}_x$ perovskites restricts MA dynamics. *J. Mater. Chem. A* **11,** 4587–4597. `http://dx.doi.org/10.1039/D2TA09069D` (9 2023).
2. Engel, E. A., Kapil, V. & Ceriotti, M. Importance of Nuclear Quantum Effects for NMR Crystallography. *J. Phys. Chem. Lett.* **12,** 7701–7707. (2024) (Aug. 2021).
3. Tsitsvero, M., Pirillo, J., Hijikata, Y. & Komatsuzaki, T. NMR Spectrum Prediction for Dynamic Molecules by Machine Learning: A Case Study of Trefoil Knot Molecule. *The Journal of Chemical Physics* **158,** 194108. ISSN: 0021-9606. (2024) (May 2023).
4. Vold, R. L. & Hoatson, G. L. Effects of jump dynamics on solid state nuclear magnetic resonance line shapes and spin relaxation times. *Journal of Magnetic Resonance* **198,** 57–72. ISSN: 1090-7807. `https://www.sciencedirect.com/science/article/pii/S1090780709000251` (2009).
5. Rhodes, B. *et al.* 17O NMR spectroscopy reveals CO2 speciation and dynamics in hydroxide-based carbon capture materials. (2024).
6. Pickard, C. J. & Mauri, F. All-electron magnetic response with pseudopotentials: NMR chemical shifts. *Physical Review B* **63,** 245101 (2001).
7. Yates, J. R., Pickard, C. J. & Mauri, F. Calculation of NMR chemical shifts for extended systems using ultrasoft pseudopotentials. *Physical Review B* **76,** 024401 (2007).
8. Bonhomme, C. *et al.* First-principles calculation of NMR parameters using the gauge including projector augmented wave method: a chemist's point of view. *Chemical Reviews* **112,** 5733–5779 (2012).
9. Valenzuela Reina, J., Civaia, F., Harper, A. F., Scheurer, C. & Köcher, S. The EFG Rosetta Stone: Translating between DFT Calculations and Solid State NMR Experiments. *Faraday Discuss.,* 10.1039.D4FD00075G. ISSN: 1359-6640, 1364-5498. (2024) (2024).
10. Behler, J. & Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **98,** 146401. (2024) (Apr. 2007).

11. Deringer, V. L. *et al.* Gaussian process regression for materials and molecules. *Chemical Reviews* **121,** 10073–10141 (2021).
12. Deringer, V. L., Caro, M. A. & Csányi, G. Machine learning interatomic potentials as emerging tools for materials science. *Advanced Materials* **31,** 1902765 (2019).
13. Pickard, C. J. Ephemeral data derived potentials for random structure search. *Physical Review B* **106,** 014102 (2022).
14. Drautz, R. Atomic cluster expansion for accurate and transferable interatomic potentials. *Physical Review B* **99,** 014104 (2019).
15. Witt, W. C. *et al.* ACEpotentials.Jl: A Julia Implementation of the Atomic Cluster Expansion. *J. Chem. Phys.* **159.** arXiv: `2309.03161 [physics]`. (2024) (Sept. 2023).
16. Batatia, I., Kovacs, D. P., Simm, G., Ortner, C. & Csanyi, G. MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields. *Advances in Neural Information Processing Systems* **35,** 11423–11436. (2023) (Dec. 2022).
17. Batzner, S. *et al.* E(3)-Equivariant Graph Neural Networks for Data-Efficient and Accurate Interatomic Potentials. *Nat Commun* **13,** 2453. ISSN: 2041-1723. (2024) (May 2022).
18. Batatia, I. *et al. A foundation model for atomistic materials chemistry* arXiv - this paper is currently a pre-print, https://arxiv.org/abs/2401.00096. 2024. arXiv: `2401.00096 [physics.chem-ph]`.
19. Kaur, H. *et al. Data-Efficient Fine-Tuning of Foundational Models for First-Principles Quality Sublimation Enthalpies* May 2024. arXiv: `2405.20217 [cond-mat, physics:physics]`. (2024).
20. Deng, B. *et al. Overcoming Systematic Softening in Universal Machine Learning Interatomic Potentials by Fine-Tuning* May 2024. arXiv: `2405.07105 [cond-mat]`. (2024).
21. Shoghi, N. *et al. From Molecules to Materials: Pre-training Large Generalizable Models for Atomic Property Prediction* May 2024. arXiv: `2310.16802 [cs]`. (2024).
22. Schaaf, L., Fako, E., De, S., Schäfer, A. & Csányi, G. *Accurate Reaction Barriers for Catalytic Pathways: An Automatic Training Protocol for Machine Learning Force Fields* Jan. 2023. arXiv: `2301.09931 [physics]`. (2023).
23. Bernstein, N., Csányi, G. & Deringer, V. L. De Novo Exploration and Self-Guided Learning of Potential-Energy Surfaces. *npj Comput Mater* **5,** 1–9. ISSN: 2057-3960. (2022) (Oct. 2019).
24. Zhang, L., Lin, D.-Y., Wang, H., Car, R. & E, W. Active Learning of Uniformly Accurate Interatomic Potentials for Materials Simulation. *Phys. Rev. Mater.* **3,** 023804. (2023) (Feb. 2019).
25. Smith, J. S., Nebgen, B., Lubbers, N., Isayev, O. & Roitberg, A. E. Less Is More: Sampling Chemical Space with Active Learning. *J. Chem. Phys.* **148,** 241733. ISSN: 0021-9606. (2022) (June 2018).
26. Goedecker, S. Minima hopping: An efficient search method for the global minimum of the potential energy surface of complex molecular systems. *The Journal of chemical physics* **120,** 9911–9917 (2004).
27. Bien, C. E. *et al.* Bioinspired Metal–Organic Framework for Trace CO2 Capture. *Journal of the American Chemical Society* **140,** 12662–12666. ISSN: 0002-7863. `https://doi.org/10.1021/jacs.8b06109` (2018).
28. Berge, A. H. *et al.* Revealing carbon capture chemistry with 17-oxygen NMR spectroscopy. *Nature Communications* **13,** 7763. ISSN: 2041-1723. `https://doi.org/10.1038/s41467-022-35254-w` (2022).
29. Harper, A. F., Köcher, S. S., Reuter, K. & Scheurer, C. *Performance Metrics for Tensorial Learning: Prediction of Li4Ti5O12 Nuclear Magnetic Resonance Observables at Experimental Accuracy* Apr. 2024. (2024).