Efficient Nudged Elastic Band Method using Neural Network Bayesian Algorithm Execution

Pranav Kakhandiki
Department of Applied Physics
Stanford University
Stanford, CA 94305
pkakhand@stanford.edu

Daniel Ratner
SLAC National Accelerator Laboratory
Menlo Park, CA, 94025
dratner@slac.stanford.edu

Sean Gasiorowski
SLAC National Accelerator Laboratory
Menlo Park, CA, 94025
sgaz@slac.stanford.edu

Abstract

Minimum energy pathways (MEPs) provide critical insights about transition states and energy barriers for chemical systems. A popular method for MEP discovery is the nudged elastic band (NEB) algorithm, which involves an expensive optimization using hundreds to tens of thousands of potentially expensive simulations. AI methods can help reduce this cost, but have historically focused on either directly running NEB on static, pre-trained models or actively updating simple (Gaussian process) simulation surrogates. To our knowledge, we are the first to unite these two regimes by using Bayesian Algorithm Execution (BAX), a technique from Bayesian experimental design, to fine-tune EquiformerV2, a foundation model. We demonstrate the resulting neural network BAX (NN-BAX) method on Lennard-Jones transitions and observe that NN-BAX requires one to two orders of magnitude fewer energy/force function evaluations compared to classical NEB, with negligible loss in accuracy of the energy barrier and transition state prediction. We highlight that this targeted fine-tuning procedure retains the simulation efficiency of previous active approaches, while allowing scalability to systems of much higher complexity and dimensionality.

1 Introduction

Minimum energy pathways (MEPs) describe the most probable paths that atomic systems follow when transitioning between stable states. MEPs are important because they dictate the mechanisms and corresponding energy barriers for transitions [1]. Nudged elastic band (NEB) is a popular method used for finding MEPs in atomic systems [2–4]. In its basic form [2], NEB begins with a set of "moving images" in atomic space between the initial and final state being studied. Force and energy calculations from physical simulations of the system are then used to nudge the band of images until a convergence criterion is met—commonly the maximum perpendicular force on any atom in any image, f_{max}, falling below a threshold. Each intermediate moving image is connected by an elastic spring force to ensure the band does not collapse to a single point. Since its development, various versions of NEB have been created to scale to more difficult systems or to reduce computational cost [5–9]. However, a major drawback of these "classical" NEB methods is that many simulations are required, namely one per image for each NEB iteration. Such simulations often involve density

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: .

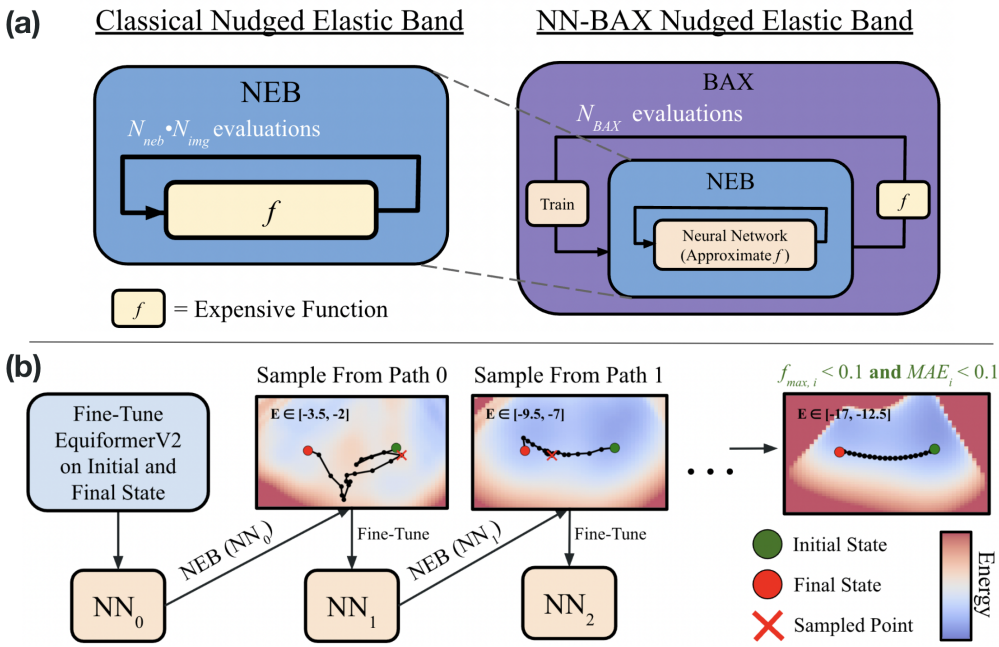

Figure 1: NN-BAX Pipeline. a) Classical NEB versus NN-BAX NEB, where the expensive simulator f is no longer called in the NEB Loop. The number of function evaluations in classical NEB scales with the number of images times the number of NEB iterations, $N_{neb} \cdot N_{img}$, whereas in NN-BAX it scales with the number of BAX iterations, N_{BAX}. b) NN-BAX Loop. The EquiformerV2 model is initialized and trained on the initial and final state. In each loop, NEB is run using the trained model as an approximation for f, a sample is acquired from the resulting path, and the sampled point is added to the training set for the subsequent iteration. Colormap energy ranges (eV) are provided.

functional theory (DFT) calculations, which can be expensive computationally. For example OC22, a large modern dataset, required >240 million core-hours to generate ~62K DFT relaxations [10].

Numerous efforts have been made to reduce the computational cost of NEB with machine learning [11–23]. In one paradigm, neural network surrogate models are trained on large, diverse datasets. These models then serve as direct calculators for downstream applications, such as NEB calculations on complex systems. An alternative approach involves sequentially updating a simulation surrogate for a system of interest using techniques from active learning or Bayesian experimental design. Bayesian experimental design techniques can be used to infer properties of black-box functions in limited function queries. Common approaches fall into the regime of Bayesian optimization (BO) [24], where the aim is to maximize or minimize the unknown function. Bayesian algorithm execution (BAX) [25] generalizes the optimization objective to any computable function property (called the BAX "algorithm"). Specifically, with BAX, we choose function queries that target the output of the NEB optimization on the unknown function defined by the physics simulator. This approach is similar to that of [12], and we emphasize the connection to the broader set of methodology.

Bayesian experimental design approaches have historically used Gaussian processes (GPs) as a surrogate model, limiting their expressivity and scalability. In our work, we aim to take advantage of the development of large neural network surrogates for materials systems and apply them within the BAX acquisition framework. We refer to this neural network BAX methodology as NN-BAX. We demonstrate NN-BAX on atomic systems governed by the Lennard-Jones (LJ) potential:

$$V_{\mathrm{LJ}}(r) = 4\varepsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right], \tag{1}$$

because of its physical relevance and its well-studied transitions [26–34]. Unlike energy/force computations involving density functional theory, the LJ potential is inexpensive to call. However, it retains much of the complexity of similar chemical systems, making it ideal for testing our method. We analyze systems with 7 and 38 identical atoms, abbreviated as LJ_7 and LJ_{38}. The dimensionality

of these systems, particularly the 114-dimensional LJ_{38}, makes GP surrogates infeasible. We use EquiformerV2 [35] as our surrogate model, pre-trained on the OMat24 [36] dataset. OMat24 does not include LJ systems, and the pre-trained EquiformerV2 model does not work well for zero-shot NEB calculations on LJ. We therefore treat it as a foundation model and tune it in a few-shot-learning manner using BAX.

2 Methods

NN-BAX Algorithm There are a variety of forms of BAX acquisition. InfoBAX [25], an information-based acquisition function, aims to maximize the expected information gain (EIG) about an algorithm output. However, it is costly to compute, involving several model retrainings to estimate the associated entropies. Heuristic or approximate BAX algorithms have been proposed that are much more efficient, such as MeanBAX [37] and PS-BAX [38] which run the algorithm on the posterior mean or a posterior sample and then take the output point with the highest predicted uncertainty, avoiding the cost of computing the full EIG. In this work, we take a similar, heuristic approach. Our BAX acquisition involves running NEB on a deterministic model of the potential energy/forces. We then randomly sample from the resulting NEB path to choose the next point to acquire and update the model for the next loop. A visualization of the pipeline is shown in Figure 1.

System Setup and Training We analyze our method on a set of transitions of LJ_7 and LJ_{38}. To select transitions, we first find local minima using basin hopping [26]. We then run classical NEB between pairs of local minima, observing that the convergence hyperparameter f_{max} varies between 0.05 and 0.3eV/Å. To check stability of the resulting MEPs, we randomly perturb the positions of the images from the converged path by $\sigma_x \in [0,\, 0.1]$ Å and re-run the NEB minimization. If after 100 iterations the path does not move and no lower energy path is discovered, the MEP is considered a "ground truth" target for our NN-BAX optimization. For both classical and NN-BAX NEB calculation, we use the Atomic Simulation Environment (ASE) library [39]. All NEB uses the string method [6], which ignores spring forces and evenly redistributes images along the path every iteration, and the FIRE optimizer [40] with parameters $dt = 0.1$ and $dt_{max} = 1.0$. In NN-BAX, we treat the number of NEB iterations, N_{neb}, as an additional hyperparameter because for earlier, less accurate models, NEB often does not converge below the given f_{max} criterion. For simplicity, we base the number of NEB iterations within NN-BAX on the number of iterations it took classical NEB to converge. In general, we would not have access to N_{neb}, but the procedure is robust to this choice, and methods such as patience [41] will further reduce sensitivity to N_{neb}.

For each NN-BAX iteration, we initialize the model to the 153 million parameter EquiformerV2 trained on the OMat24 dataset. EquiformerV2 combines the Transformer architecture [42] with consideration of relevant symmetries for materials systems: atomic energies are modeled as invariant under global translations and rotations of the system, while atomic forces are modeled as equivariant, transforming consistently with rotations and translations of the atomic positions. EquiformerV2 has demonstrated strong performance in benchmarks [43] and previous non active learning (zero-shot) neural network NEB studies such as CatTsunami [11]. The inputs to the network are the positions of the atoms in a given LJ configuration, x_i, and the outputs are the forces on each atom and the potential energy of the system. Our version of EquiformerV2 also provides stress predictions, but these are irrelevant for our systems, so we initialize all stresses to zero and ignore them.

3 Results and Discussion

We analyze four transitions of varying numbers of intermediate minima, corresponding to increasing difficulty. For LJ_7 we look at paths with 0 and 3 intermediate minima, and for LJ_{38} we analyze paths with 0 and 1 intermediate minima. These paths converged for classical NEB with f_{max} values of 0.05, 0.15, 0.05, and 0.3eV/Å respectively, and passed the stability check. Figure 2 compares the converged NN-BAX paths with their classical NEB counterparts. The visuals on top display the atomic structures of the initial and final states. The upper plots show energy versus image index for both classical NEB and NN-BAX, with very close agreement between the two results. Of the four, the worst predicted energy barrier is 0.6% greater than the ground truth. The bottom row shows a two dimensional principal component analysis (PCA) projection of the MEPs and potential energy landscapes, with projection computed from the ground truth path in atomic coordinate space. We

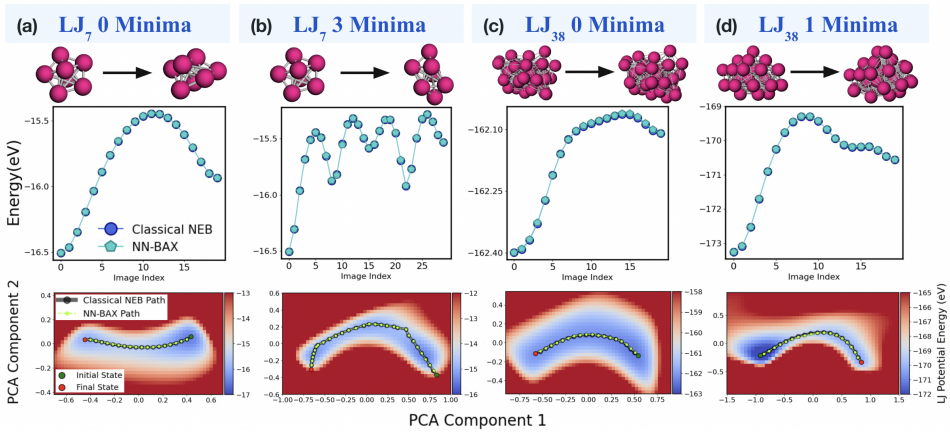

Figure 2: Results from running NN-BAX on various paths with LJ_7 and LJ_{38} systems. The upper row compares the energy profile predicted by NN-BAX to classical NEB, along with atomic visuals for the transition. The bottom row shows dimension-reduced principal component analysis plots of the energy, with the NN-BAX and classical NEB paths overlaid.

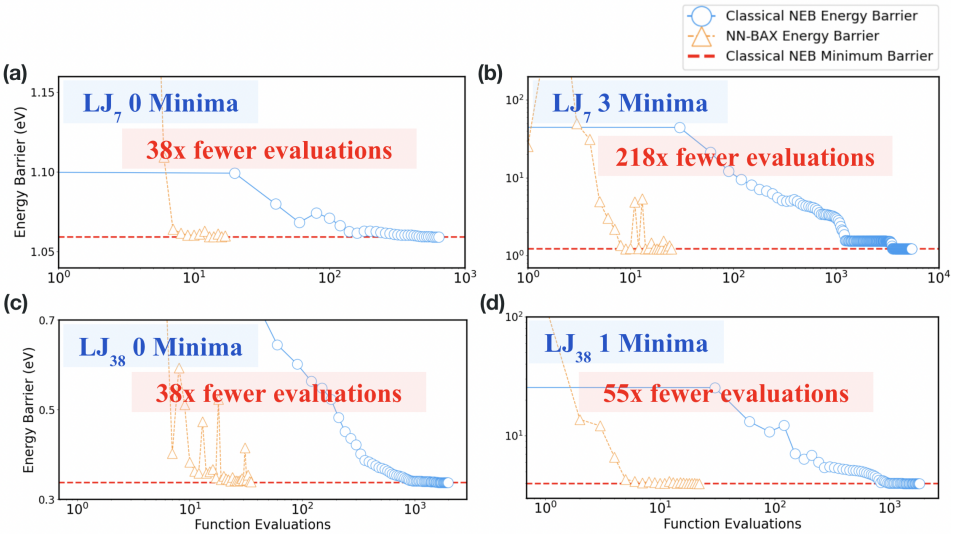

Figure 3: Energy barrier v. function evaluations, for various LJ_7 and LJ_{38} systems

observe that the NN-BAX paths qualitatively line up with the classical NEB paths, indicating that NN-BAX is finding the same structures as classical NEB.

Figure 3 displays the speed-up NN-BAX achieves with respect to classical NEB, in terms of function evaluations. As the LJ potential is quick to call, in this setting there is significant overhead from the NN-BAX procedure. However, for systems running expensive DFT simulations, this becomes negligible, making number of function calls an appropriate figure of merit. We achieve a 1-2 orders of magnitude reduction in function calls across all four paths. In addition to the f_{max} criterion used by classical NEB, we introduce a metric MAE_i, defined as the mean absolute error of model $i-1$ computed on sampled point i, to define convergence of the full NN-BAX procedure. Reduction in MAE_i indicates that the model is performing better on points not in the training set and, once NN-BAX begins predicting similar final paths, the sampled points grow increasingly similar, an easier modeling task. For NN-BAX to converge, we therefore require NEB to converge (f^i_{max} below threshold) and MAE_i to fall below a threshold value. Some energy barrier values for NN-BAX are negligibly lower than the ground truth—this happens because of the discrete nature of the path, resulting in images not sitting precisely at the MEP peak. Finally we observe that the LJ_7 path with 3 minima achieves a much greater speedup than the other paths. We hypothesize this occurs due to

the high number of NEB iterations required for classical NEB to converge, despite the energy/force landscape modeling difficulty being similar to the other paths. Thus because the number of function evaluations in NN-BAX no longer scales with N_{neb}, we observe a larger speedup for this path.

4 Conclusion

The nudged elastic band algorithm is critical in finding transition states and energy barriers in chemical systems. Lennard Jones systems hold much of the complexity of the broader group of chemical systems of interest, and allow for analysis of high dimensional atomic systems. In this paper, we unified efforts in active learning and foundation models using BAX. Specifically, we applied NN-BAX to NEB on Lennard Jones systems and observed a speedup of 1-2 orders of magnitude with respect to number of function evaluations. We emphasize that NN-BAX describes a broader system-agnostic procedure, and therefore we expect our work to extend to both new models and NEB applications, including ones that require expensive simulations like DFT.

Acknowledgments and Disclosure of Funding

We acknowledge Sathya Chitturi and Guanzhi Li for helpful discussions. SG and PK were supported by the Department of Energy, Laboratory Directed Research and Development program at SLAC National Accelerator Laboratory, under contract DE-AC02-76SF00515.

References

[1] Henry Eyring. The activated complex in chemical reactions. *The Journal of chemical physics*, 3 (2):107–115, 1935.

[2] Hannes Jónsson, Greg Mills, and Karsten W Jacobsen. Nudged elastic band method for finding minimum energy paths of transitions. In *Classical and quantum dynamics in condensed phase simulations*, pages 385–404. World Scientific, 1998.

[3] Graeme Henkelman and Hannes Jónsson. Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *The Journal of chemical physics*, 113(22):9978–9985, 2000.

[4] Graeme Henkelman, Blas P Uberuaga, and Hannes Jónsson. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *The Journal of chemical physics*, 113(22):9901–9904, 2000.

[5] Semen A Trygubenko and David J Wales. A doubly nudged elastic band method for finding transition states. *The Journal of chemical physics*, 120(5):2082–2094, 2004.

[6] E Weinan, Weiqing Ren, and Eric Vanden-Eijnden. String method for the study of rare events. *Physical Review B*, 66(5):052301, 2002.

[7] P Maragakis, Stefan A Andreev, Yisroel Brumer, David R Reichman, and Efthimios Kaxiras. Adaptive nudged elastic band approach for transition state calculation. *The Journal of chemical physics*, 117(10):4651–4658, 2002.

[8] Daniel Sheppard, Penghao Xiao, William Chemelewski, Duane D Johnson, and Graeme Henkelman. A generalized solid-state nudged elastic band method. *The Journal of chemical physics*, 136(7), 2012.

[9] Esben L Kolsbjerg, Michael N Groves, and Bjørk Hammer. An automated nudged elastic band method. *The Journal of chemical physics*, 145(9), 2016.

[10] Richard Tran, Janice Lan, Muhammed Shuaibi, Brandon M Wood, Siddharth Goyal, Abhishek Das, Javier Heras-Domingo, Adeesh Kolluru, Ammar Rizvi, Nima Shoghi, et al. The open catalyst 2022 (oc22) dataset and challenges for oxide electrocatalysts. *ACS Catalysis*, 13(5): 3066–3084, 2023.

[11] Brook Wander, Muhammed Shuaibi, John R Kitchin, Zachary W Ulissi, and C Lawrence Zitnick. Cattsunami: Accelerating transition state energy calculations with pretrained graph neural networks. *ACS Catalysis*, 15(7):5283–5294, 2025.

[12] José A Garrido Torres, Paul C Jennings, Martin H Hansen, Jacob R Boes, and Thomas Bligaard. Low-scaling algorithm for nudged elastic band calculations using a surrogate machine learning model. *Physical review letters*, 122(15):156001, 2019.

[13] Olli-Pekka Koistinen, Vilhjálmur Ásgeirsson, Aki Vehtari, and Hannes Jónsson. Nudged elastic band calculations accelerated with gaussian process regression based on inverse interatomic distances. *Journal of chemical theory and computation*, 15(12):6738–6751, 2019.

[14] Xiang Fu, Brandon M Wood, Luis Barroso-Luque, Daniel S Levine, Meng Gao, Misko Dzamba, and C Lawrence Zitnick. Learning smooth and expressive interatomic potentials for physical property prediction. *arXiv preprint arXiv:2502.12147*, 2025.

[15] Jun Luo, Tao Fan, Jiawei Zhang, Pengfei Qiu, Xun Shi, and Lidong Chen. Automatic identification of slip pathways in ductile inorganic materials by combining the active learning strategy and neb method. *npj Computational Materials*, 11(1):41, 2025.

[16] Bowen Deng, Yunyeong Choi, Peichen Zhong, Janosh Riebesell, Shashwat Anand, Zhuohan Li, KyuJung Jun, Kristin A Persson, and Gerbrand Ceder. Systematic softening in universal machine learning interatomic potentials. *npj Computational Materials*, 11(1):9, 2025.

[17] Bowen Li, Jin Xiao, Ya Gao, John ZH Zhang, and Tong Zhu. Transition state searching accelerated by neural network potential. *Journal of Chemical Information and Modeling*, 65(5): 2297–2303, 2025.

[18] Simone Perego and Luigi Bonati. Data efficient machine learning potentials for modeling catalytic reactivity via active learning and enhanced sampling. *npj Computational Materials*, 10 (1):291, 2024.

[19] Lars L Schaaf, Edvin Fako, Sandip De, Ansgar Schäfer, and Gábor Csányi. Accurate energy barriers for catalytic reaction pathways: an automatic training protocol for machine learning force fields. *npj Computational Materials*, 9(1):180, 2023.

[20] Ralf Meyer, Klemens S Schmuck, and Andreas W Hauser. Machine learning in computational chemistry: An evaluation of method performance for nudged elastic band calculations. *Journal of Chemical Theory and Computation*, 15(11):6513–6523, 2019.

[21] Olli-Pekka Koistinen, Freyja B Dagbjartsdóttir, Vilhjálmur Ásgeirsson, Aki Vehtari, and Hannes Jónsson. Nudged elastic band calculations accelerated with gaussian process regression. *The Journal of chemical physics*, 147(15), 2017.

[22] Mathias Schreiner, Arghya Bhowmik, Tejs Vegge, Peter Bjørn Jørgensen, and Ole Winther. Neuralneb—neural networks can find reaction paths fast. *Machine Learning: Science and Technology*, 3(4):045022, 2022.

[23] Chong Teng, Yang Wang, and Junwei Lucas Bao. Physical prior mean function-driven gaussian processes search for minimum-energy reaction paths with a climbing-image nudged elastic band: a general method for gas-phase, interfacial, and bulk-phase reactions. *Journal of Chemical Theory and Computation*, 20(10):4308–4324, 2024.

[24] Donald R Jones, Matthias Schonlau, and William J Welch. Efficient global optimization of expensive black-box functions. *Journal of Global optimization*, 13(4):455–492, 1998.

[25] Willie Neiswanger, Ke Alexander Wang, and Stefano Ermon. Bayesian algorithm execution: Estimating computable properties of black-box functions using mutual information. In *International Conference on Machine Learning*, pages 8005–8015. PMLR, 2021.

[26] David J Wales and Jonathan PK Doye. Global optimization by basin-hopping and the lowest energy structures of lennard-jones clusters containing up to 110 atoms. *The Journal of Physical Chemistry A*, 101(28):5111–5116, 1997.

[27] Jonathan PK Doye and David J Wales. Saddle points and dynamics of lennard-jones clusters, solids, and supercooled liquids. *The Journal of Chemical Physics*, 116(9):3777–3788, 2002.

[28] Peter Schwerdtfeger and David J Wales. 100 years of the lennard-jones potential. *Journal of Chemical Theory and Computation*, 20(9):3379–3405, 2024.

[29] David J Wales. Rearrangements of 55-atom lennard-jones and (c60) 55 clusters. *The Journal of chemical physics*, 101(5):3750–3762, 1994.

[30] Jonathan PK Doye, David J Wales, and Mark A Miller. Thermodynamics and the global optimization of lennard-jones clusters. *The Journal of Chemical Physics*, 109(19):8143–8153, 1998.

[31] Jonathan PK Doye, Mark A Miller, and David J Wales. The double-funnel energy landscape of the 38-atom lennard-jones cluster. *The Journal of Chemical Physics*, 110(14):6896–6906, 1999.

[32] Ray M Sehgal, Dimitrios Maroudas, and David M Ford. Phase behavior of the 38-atom lennard-jones cluster. *The Journal of Chemical Physics*, 140(10), 2014.

[33] David J Wales, Mark A Miller, and Tiffany R Walsh. Archetypal energy landscapes. *Nature*, 394(6695):758–760, 1998.

[34] Matthias U Bohner, Jan Meisner, and Johannes Kastner. A quadratically-converging nudged elastic band optimizer. *Journal of Chemical Theory and Computation*, 9(8):3498–3504, 2013.

[35] Yi-Lun Liao, Brandon Wood, Abhishek Das, and Tess Smidt. Equiformerv2: Improved equivariant transformer for scaling to higher-degree representations. *arXiv preprint arXiv:2306.12059*, 2023.

[36] Luis Barroso-Luque, Muhammed Shuaibi, Xiang Fu, Brandon M Wood, Misko Dzamba, Meng Gao, Ammar Rizvi, C Lawrence Zitnick, and Zachary W Ulissi. Open materials 2024 (omat24) inorganic materials dataset and models. *arXiv preprint arXiv:2410.12771*, 2024.

[37] Sathya R Chitturi, Akash Ramdas, Yue Wu, Brian Rohr, Stefano Ermon, Jennifer Dionne, Felipe H da Jornada, Mike Dunne, Christopher Tassone, Willie Neiswanger, et al. Targeted materials discovery using bayesian algorithm execution. *npj Computational Materials*, 10(1):156, 2024.

[38] Chu Xin Cheng, Raul Astudillo, Thomas A Desautels, and Yisong Yue. Practical bayesian algorithm execution via posterior sampling. *Advances in Neural Information Processing Systems*, 37:135186–135207, 2024.

[39] Ask Hjorth Larsen, Jens Jørgen Mortensen, Jakob Blomqvist, Ivano E Castelli, Rune Christensen, Marcin Dułak, Jesper Friis, Michael N Groves, Bjørk Hammer, Cory Hargus, et al. The atomic simulation environment—a python library for working with atoms. *Journal of Physics: Condensed Matter*, 29(27):273002, 2017.

[40] Erik Bitzek, Pekka Koskinen, Franz Gähler, Michael Moseler, and Peter Gumbsch. Structural relaxation made simple. *Physical review letters*, 97(17):170201, 2006.

[41] Lutz Prechelt. Early stopping-but when? In *Neural Networks: Tricks of the trade*, pages 55–69. Springer, 2002.

[42] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[43] Janosh Riebesell, Rhys EA Goodall, Philipp Benner, Yuan Chiang, Bowen Deng, Alpha A Lee, Anubhav Jain, and Kristin A Persson. Matbench discovery–a framework to evaluate machine learning crystal stability predictions. *arXiv preprint arXiv:2308.14920*, 2023.

[44] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

5 Supplementary Materials

5.1 Transition State Predictions

A state along the reaction path that corresponds to a maximum of potential energy is called a transition state. A transition state is often of particular interest because it dictates the activation energy of the reaction and corresponds to bonds breaking and new bonds forming [1]. Transition states are helpful in figuring out reaction rates in chemical systems and in determining whether reactions are physically possible. In Figure 4 we display the predicted structures for the highest energy transition state of each reaction. We observe the NN-BAX structures qualitatively match the classical NEB structures.

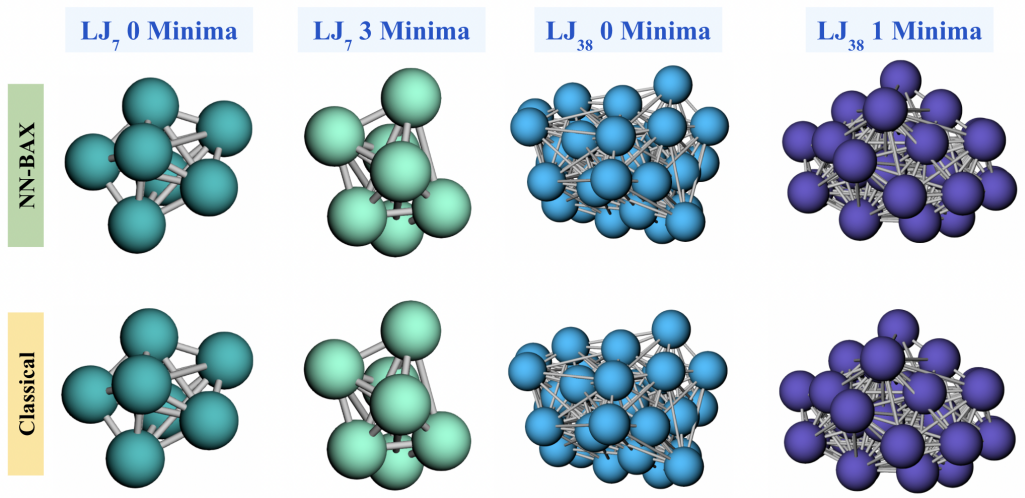

Figure 4: Transition state predictions. The NN-BAX predictions and the classical ground truth atomic arrangements are displayed.

5.2 NN-BAX Convergence

In this paper we introduce a new convergence hyperparameter, MAE_i, which corresponds to the error in the prediction of model $i-1$ on datapoint i. In Figure 5 MAE_i is plotted for each BAX iteration, for the four paths we analyze. We observe it to be a relatively stable parameter, plateauing as we increase the BAX iteration. In order for NN-BAX to converge, $\mathrm{MAE}_i < m$ and $f^i_{max,BAX} < t$ must simultaneously be satisfied, where m and t are the respective convergence thresholds. For all results presented here, we require $\mathrm{MAE}_i < 0.1$. For force convergence, we set $f_{\mathrm{max,BAX}} = n\, f_{\mathrm{max,classical}}$, where $n > 1$ allows the model predictions slightly more error than the true potential. n is a hyperparameter, so various values may work for different paths. In practice, we found $n = 2$ to work well for our paths. For further confidence in convergence we introduce a patience metric p, where the convergence criteria must be achieved p times for NN-BAX to converge. In this paper we use $p = 2$ for both LJ_7 paths and the LJ_{38} path with one minimum, and $p = 3$ for the LJ_{38} path with zero minima.

Finally we comment on convergence time and BAX overhead. Specifically the LJ_{38} 0 minima path took NN-BAX 9.3 hours to run, for 30 BAX iterations, with 200 NEB steps, and with each model trained for 50 epochs. All code was run on a single NVIDIA Tesla A100 GPU. Since calls to the LJ potential are practically instant, this is a good approximation for solely the computational overhead cost of NN-BAX. Thus for a setting in which DFT is being used, for NN-BAX to achieve a wall-clock speedup relative to Classical NEB, the DFT simulations would have to be longer than 25.4 seconds. Finally we note that the NEB calculations with the model were not parallelized, but could be, which would further improve the speedup.

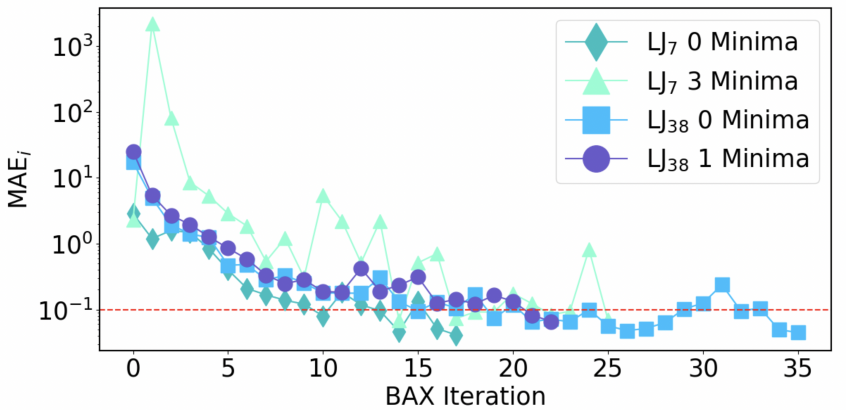

Figure 5: MAE_i convergence metric versus BAX iteration. The red line denotes the convergence threshold of 0.1

5.3 Training

In each BAX iteration, the network was trained for 50 epochs with a batch size of 2 using the AdamW optimizer [44] (weight decay 0.001, gradient clipping at 100) and an initial learning rate of 2×10^{-4}. A cosine learning rate schedule was employed via `LambdaLR`, with a warmup phase lasting 1 epoch (warmup factor 0.2) and a minimum learning rate set to 1% of the initial value. Exponential moving average (EMA) updates with a decay of 0.999 were applied throughout training. We use the mean absolute error (MAE) loss, giving forces four times the weight of energy since forces are critical in NEB precision and convergence. All models are trained on a single NVIDIA Tesla A100 GPU at the SLAC Shared Science Data Facility.