

UC Berkeley

UC Berkeley Previously Published Works

Title

Analytical ab initio hessian from a deep learning potential for transition state optimization.

Permalink

<https://escholarship.org/uc/item/8n25f27s>

Journal

Nature Communications, 15(1)

Authors

Yuan, Eric

Kumar, Anup

Guan, Xingyi

et al.

Publication Date

2024-10-14

DOI

10.1038/s41467-024-52481-5

Peer reviewed

Analytical ab initio hessian from a deep learning potential for transition state optimization

Received: 28 March 2024

Accepted: 6 September 2024

Published online: 14 October 2024

Check for updates

Eric C.-Y. Yuan^{1,2,9}, Anup Kumar^{3,9}, Xingyi Guan^{1,2}, Eric D. Hermes⁴, Andrew S. Rosen^{5,6}, Judit Zádor⁷, Teresa Head-Gordon^{1,2,8} ✉ & Samuel M. Blau³ ✉

Identifying transition states—saddle points on the potential energy surface connecting reactant and product minima—is central to predicting kinetic barriers and understanding chemical reaction mechanisms. In this work, we train a fully differentiable equivariant neural network potential, NewtonNet, on thousands of organic reactions and derive the analytical Hessians. By reducing the computational cost by several orders of magnitude relative to the density functional theory (DFT) ab initio source, we can afford to use the learned Hessians at every step for the saddle point optimizations. We show that the full machine learned (ML) Hessian robustly finds the transition states of 240 unseen organic reactions, even when the quality of the initial guess structures are degraded, while reducing the number of optimization steps to convergence by 2–3× compared to the quasi-Newton DFT and ML methods. All data generation, NewtonNet model, and ML transition state finding methods are available in an automated workflow.

Computational identification of transition states (TSs) on the quantum mechanical potential energy surface (PES) is central to predicting reaction barriers and understanding chemical reactivity^{1,2}. The height of the barrier exponentially impacts the reaction rate coefficient via the Eyring equation, and the geometric character of the metastable state is informative about the kinetic mechanism, making TSs key to describing a broad range of chemical kinetic outcomes for enzymes, next-generation synthetic catalysts, batteries, and conformational changes of molecules and materials^{3,4}.

Transition states are first-order saddle points, and locating them via mathematical optimization is particularly challenging on high dimensional PESs relevant in complex molecular systems. Locating an equilibrium geometry, i.e., a local minimum on the PES, can be found using bracketing methods based on function evaluations (0th order)^{5,6}

and with methods that use gradient information, such as steepest descent⁷ or conjugate gradient⁸ (1st order). Under the quadratic approximation, identification of the local minima can be more robustly achieved in fewer steps by 2nd order methods⁹ using the Hessian matrix, whose elements H_{ij} are defined as the second derivative of the energy E with respect to atomic positions R_i and R_j . However, a metastable first-order saddle point is characterized by a single negative Hessian eigenvalue, and hence 2nd order methods are indispensable to optimize for molecular TS energies and geometries¹. The Newton-Raphson (NR) method and its variants, including restricted and augmented methods like the trust radius method (TRM) and the rational function optimization (RFO) method, select the displacement vector $\Delta R^{(k)}$ (or in internal coordinates) at step k using the inverse Hessian $[H^{(k)}]^{-1}$ and the gradient $g^{(k)}$ ^{10–12}. The geometry direct inversion of the

¹Kenneth S. Pitzer Theory Center and Department of Chemistry, University of California, Berkeley, CA, USA. ²Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. ³Energy Technologies Area, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. ⁴Quantum-Si, Branford, CT, USA. ⁵Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. ⁶Department of Materials Science and Engineering, University of California, Berkeley, CA, USA. ⁷Combustion Research Facility, Sandia National Laboratories, Livermore, CA, USA. ⁸Departments of Bioengineering and Chemical and Biomolecular Engineering, University of California, Berkeley, CA, USA. ⁹These authors contributed equally: Eric C.-Y. Yuan, Anup Kumar.

✉ e-mail: thg@berkeley.edu; smblau@lbl.gov

iterative subspaces (GDIIIS) method and its variants similarly utilize Hessians and gradients to define a search space for optimization¹³.

However, the evaluation of analytical Hessians for ab initio methods such as density functional theory (DFT) requires solving coupled-perturbed equations, which scale one power of system size N higher than the energy or the gradient and thus can be prohibitively expensive. Consequently, almost all TS optimization approaches rely on constructing cheaper approximate Hessians using only gradient information to avoid expensive Hessian calculations, in general referred to as quasi-Newton (QN) methods^{14–20}. Arguably the most widely used Hessian approximation for minimization is the Broyden–Fletcher–Goldfarb–Shanno (BFGS) method, where the Hessian is iteratively updated using a rank-2 matrix generated from the displacements and gradients. Such Hessian updates, however, are positive definite by design and thus cannot be applied to TS searches. Instead, methods such as symmetric rank-one (SRI) or Murtagh–Sargent, Powell–symmetric-Broyden (PSB), Murtagh–Sargent–Powell (MSP), Bofill, and TS-BFGS methods are developed for an indefinite approximate Hessian^{21–25}. On a complex PES, an optimization step can displace the molecule from the preceding quadratic region such that the resulting updated approximate Hessian quickly diverges from the true one and thus requires expensive reconstruction¹². Even though double- and single-ended interpolation methods such as nudged elastic band (NEB)^{26,27}, quadratic synchronous transit (QST)²⁸, and growing string method (GSM)²⁹ have been well-established in the field in the past few decades, a subsequent TS optimization with QN updates is often integrated into the workflow. Despite all these efforts, TS optimization still requires significant user involvement and relies on trial and error when robust Hessian information is absent. However, if the full Hessian is available at every optimization step, concerns regarding the quality of the initial Hessian and subsequent updates become much less of a problem when determining a TS.

The recent development of deep learning models for the PES provides an alternative possibility for acquiring and applying the Hessian in chemically relevant tasks^{30–34}. Intuitively, the power of a fully differentiable machine learning (ML) force field does not stop at forces or gradients but also broadly applies to second (and higher order) derivative properties such as the Hessian matrix H_{ij} . In this case, it is possible to calculate Hessians analytically by automatic differentiation, by finite differences using gradients from the machine learning model, or by estimating Hessians using first order information as per the Davidson procedure¹. For example, such an idea has recently been explored using Gaussian process regression, where an ML PES was locally trained on semiempirical energies, forces, and optionally Hessians and used to estimate the updated Hessians^{35,36}. Yet, the high memory demand using kernel-based methods can significantly reduce the applicability on all but small systems, and the semi-empirical level of theory can be deficient for reliable chemistry.

In this work, we fine-tune an equivariant message-passing neural network (eMPNN), NewtonNet³², on an augmented version of the Transition-IX (TIX) dataset³⁷, a benchmark dataset containing ~10 million configurations generated by the NEB method on ~10 thousand gas-phase organic reactions evaluated with DFT. Although the full training data is comprised of only energies and gradients of the molecular configurations, with no Hessians provided, the whole neural network is fully differentiable such that we can infer the Hessian H_{ij} through back propagation. We then apply the ML Hessians to TS optimization on an independent data set of 240 organic reactions previously proposed by Hermes and co-workers^{38,39}, and which are outside of the training set. We have adapted the Sella code³⁹ to read in full ML Hessians to perform TS optimizations for these reactions and utilize the same code and optimization settings in order to compare against QN Hessian optimization with either ML or DFT.

We find that incorporating explicit Hessians from the NewtonNet ML model into TS optimization yields a 2–3 × reduction in search steps compared to approximate Hessian methods, demonstrating a remarkable efficiency improvement by ensuring higher-confidence search directions that are closer to the optimal path. The more accurate description of the Hessian also leads to improved robustness against structural perturbation such that the TS optimization is less reliant on a good initial guess. With our deep learning model, the Hessian calculation is over 1000× faster than the corresponding ab initio calculation and is consistently more robust in finding TSs than QN methods using the ML or DFT PES. The combination of greater efficiency, reduced reliance on good initial guesses, and robust TS convergence for unseen reactions opens opportunities to utilize full Hessians for TS optimizations with appropriately constructed data sets of complex reactive chemistry.

Results

Machine learned prediction of DFT Hessians

Figure 1a shows the NewtonNet eMPNN model in which the DFT-computed molecular energy E is predicted by transforming and aggregating atomic features a_i that accumulate local chemical environmental information from spatial neighbors a_j and interatomic distances R_{ij} through message passing layers³². The molecular energy E is then differentiated with respect to the atomic positions R_i to predict atomic forces F_i or gradients g_i , but of relevance here is that it can be auto-differentiated twice to obtain H_{ij} . We have demonstrated that the energies and forces can be predicted with excellent accuracy across a whole range of chemistry including small organic molecules³² as well as for methane and hydrogen combustion, even with a limited amount of training examples^{32,40}.

Like all ML potentials, the quality of the learned PES and its derivative properties depends on the availability of relevant training data. Our ML model is pre-trained on the ANI-1 dataset, which contains more than 20 million off-equilibrium conformations of small organic molecules up to 8 heavy atoms and is evaluated with the ω B97X density functional⁴¹ and 6-31G* basis set⁴². Figure 1b demonstrates that the original ANI-1 dataset is mostly composed of near-equilibrium geometries^{42,43} and that the reaction pathways are notably under-sampled around the metastable states of the reactions³⁷. As a result, the pre-trained ML model predicts the energies and forces accurately (with respect to the underlying DFT data) at the reactant and product states but fails significantly around the TS (Fig. 1c, d).

Hence, it is fortunate to have the TIX dataset³⁷, which is a benchmark for TS-related ML tasks, containing 9,644,740 molecular configurations generated by NEB from 10,073 organic reactions, at a level of DFT commensurate with the ANI-1 data. This data better represents the entire reaction pathway as seen in Fig. 1b and allows us to fine-tune the pre-trained model. The fine-tuned model predicts both energies and forces an order of magnitude more accurately around the TS, shown in Fig. 1c, d and SI. When using the fine-tuned model, the reaction barrier is also more smoothly interpolated between NEB images, the false identification of an energy maximum has been eliminated, and the atomic forces are correctly predicted to be negligible as expected of a first-order saddle point.

Due to the high cost of the Hessian calculation and storage, the training datasets we use do not contain ab initio Hessian reference samples. Despite the lack of such training examples, a prediction of the atomic forces from an ML model that is continuous and smooth strongly suggests the possibility of achieving Hessian predictions without explicit training on such tasks. Based on this assumption, one approach is a finite-difference Hessian estimation that can be easily realized using the gradients predicted by our model by stepping along each Cartesian axis. However, an analytical gradient of the first derivative is more cost-effective than a finite-difference method, and such a gradient can be performed as long as the neural network is at least

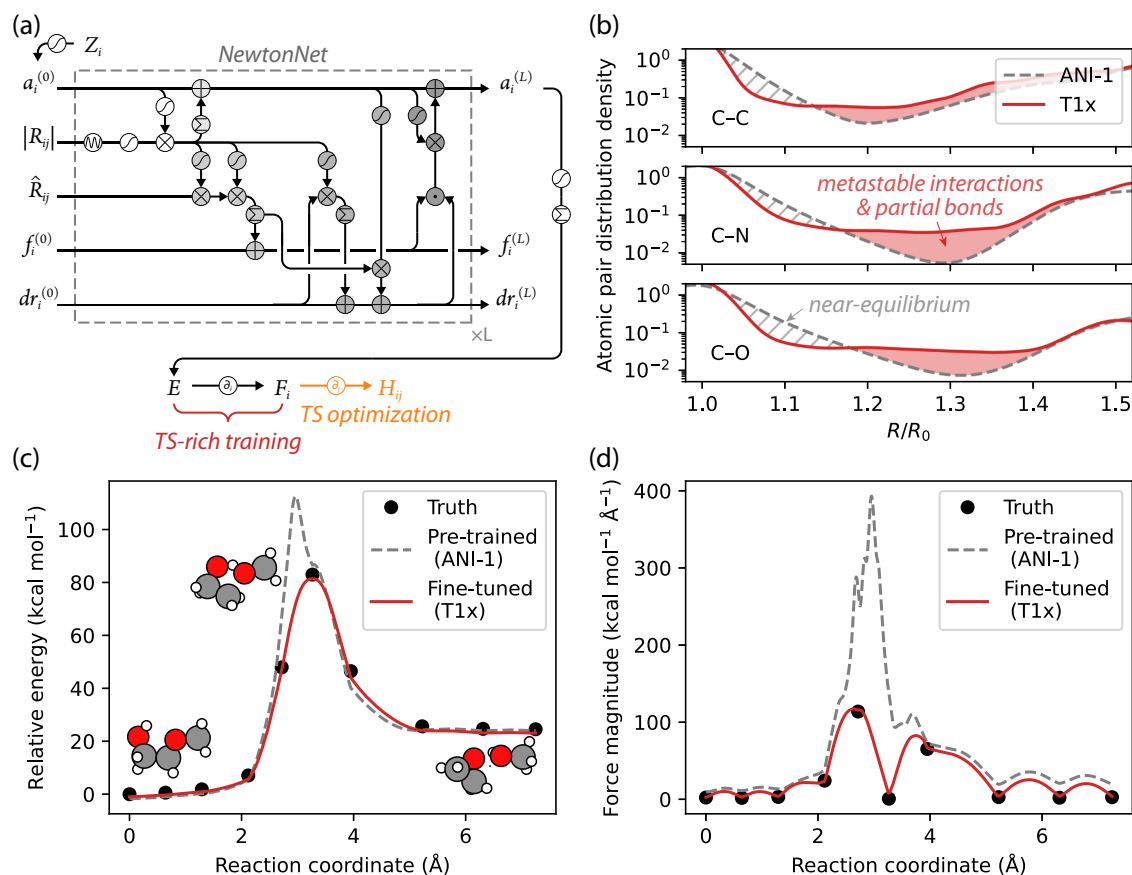


Fig. 1 | The NewtonNet model and its performance on the ANI-1 and T1x data sets. **a** The equivariant message-passing neural network designed for 3D molecular graphs with nodes $\{Z_i\}$ and edges $\{R_{ij}\}$ to predict molecular energies E and atomic forces F_i^{32} . In this work, we further differentiate the network to derive H_{ij} for TS optimization tasks. **b** The distribution of atomic pairwise distances, R , relative to equilibrium bond distances, R_0 , among datasets we used for training, where the T1x data set provides more data in the TS region, and further augmented with the ANI-1x data to add data corresponding to bond compression. The predicted (c)

potential energies and (d) forces along the reaction coordinate for an unseen reaction for the pre-trained and fine-tuned model. A comprehensive statistical analysis of energy and force prediction errors along the reaction coordinates for 1248 unseen test reactions is summarized in Supplementary Fig. S1 for the pre-trained and fine-tuned models. Details of the training protocols are described in Methods and Supplementary Fig. S2 and S3. TS: transition state. Source data for this figure are provided with this paper.

twice differentiable. In this regard, the NewtonNet model is designed using sigmoid linear unit (SiLU)⁴⁴ and a polynomial cutoff function⁴⁵ and is therefore C^2 continuous. Utilizing the automatic differentiation in our neural network, the forces and Hessians can be analytically acquired by back propagation.

Leveraging the smoothness of the deep learning PES after fine-tuning, Fig. 2 shows that reasonably accurate Hessian predictions of the reference DFT model can be acquired on molecular TSs when compared to the pre-trained model. The Hessian prediction is quantitatively accurate for both negative to positive eigenvalue regions, with eigenvalue root mean square error (RMSE) of $318 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ and eigenvector mean cosine similarity (MCS) 0.828 for TSs in unseen test reactions, improving dramatically after fine-tuning the ML model with the T1x dataset from Fig. 2a, b. It is worth noting that the majority of the error arises from the positive eigenspace with a constant 20% underestimation of Hessian eigenvalues from DFT. This can be understood in part from Fig. 1b that the T1x dataset we use for training is biased toward weaker bonds and greater anharmonicity, such that lower apparent bond strength and force constants will be observed and modeled. We therefore augmented the T1x dataset by selecting 1,232,469 molecules from the ANI-1x dataset⁴³ that share the same chemical formula with the T1x dataset but that exhibit compressed chemical bonds, which further improves accuracy of the Hessian predictions to $267 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ eigenvalue RMSE and 0.839 eigenvector

MCS in Fig. 2c. We observe that this improvement is not solely attributed to increased data volume. It effectively mitigates the underprediction of positive eigenvalues by establishing a more balanced dataset. Most importantly, the predicted Hessians by our fine-tuned model using the augmented T1x data have very accurate leftmost eigenvalue and eigenvector, which are the most critical ingredients in TS optimization and iterative Hessian diagonalization^{12,46}.

Transition state optimization using machine learned Hessians

The fine-tuned NewtonNet model for predicting TS properties is subsequently employed in practical TS optimization scenarios involving new reactions independent of the augmented ANI-1x/T1x training and test data. These include hydrogen migration reactions, endo- and exocyclization, generalized Korcek step 2 reactions, retro-ene reactions, and reverse 1,2 and 1,3 insertions (see source data); given the training data these involve only closed-shell molecules. We focus on TS optimization for these unseen reactions in order to compare a traditional QN method that approximates Hessians using gradient information from DFT calculations or ML predictions versus a full explicit ML Hessian used at every step.

We interfaced our fine-tuned NewtonNet model with Sella, a state-of-the-art open-source TS geometry optimizer³⁹. In Sella, the inter-conversion between the Cartesian coordinates and the redundant

internal coordinates is automatically handled, and the Hessian is iteratively diagonalized for the leftmost eigenvector¹² used in the geodesic saddle point optimization⁴⁷. In order to start the TS optimization for the 240 Sella benchmark reactions, we generated initial guesses with KinBot using reaction templates³⁸, where each template also defines the intended reactant and product end states for a given reaction. We employed restricted step partitioned rational function optimization (RS-PRFO)^{48–50} for the TS optimizations, with dynamically adjusted step sizes determined by evaluating the confidence of each step (see Methods for details). After TS optimization, we follow the intrinsic reaction coordinate (IRC) from the optimized TS structure to find the minimum energy path that connects the reactant and product; the robustness of the TS optimization methods is quantified by comparing the intended reactions and the predicted reactions. The complete list of found transition states of the 240 predicted benchmark reactions is summarized in Supplementary Table S1.

Figure 3 shows how optimization efficiency is dramatically improved by providing full explicit Hessians at every optimization step, which is now affordable relative to DFT as illustrated in Supplementary Fig. S9. In Fig. 3a we find that the number of steps required to converge to a TS can be reduced by 2× of that required by the QN approach using the ML (or DFT; see Supplementary Fig. S10). The trend is notably non-linear, and full-Hessian optimization is even more advantageous for challenging tasks that require larger numbers of optimization steps. If the iterative diagonalization steps for initial Hessian construction and Hessian reconstruction when the QN approximation breaks down are included, a reduction of close to 3× of the required steps is observed when considering these gradient calls. We also observed that in some most difficult cases for TS optimization with QN Hessians that take >80 steps, the optimization steps taken by full Hessians are even fewer than those with QN steps <80, which initially seemed counter-intuitive. This behavior not only

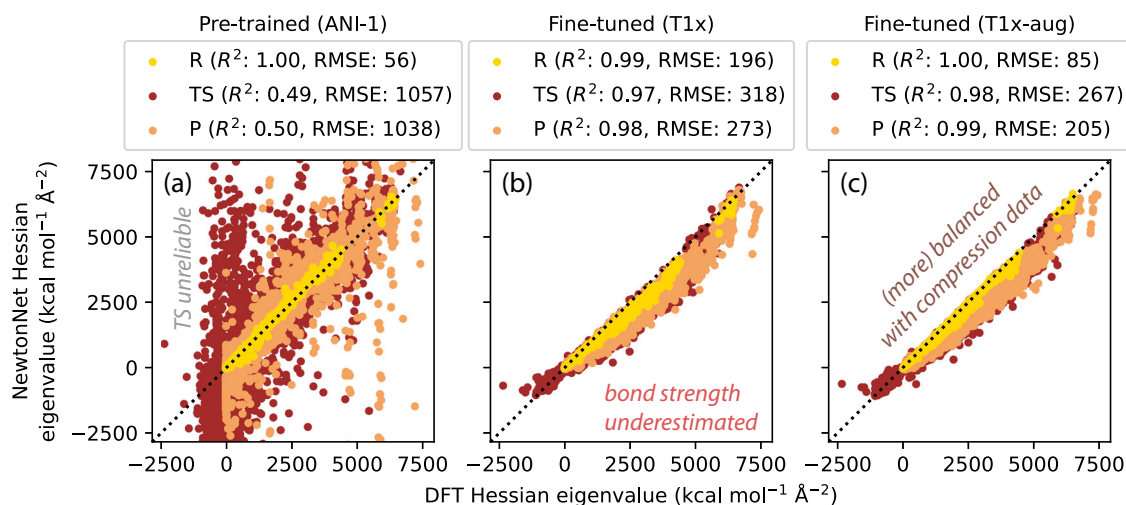


Fig. 2 | Pre-trained and fine-tuned NewtonNet performance on Hessian prediction of the test set. **a** The pre-trained model accurately predicts Hessians at R and P minima geometries but fails dramatically at TSs. **b** The fine-tuned model using the T1x data significantly improves the accuracy at TSs but with notable underestimation of Hessian eigenvalues. **c** Augmenting the T1x dataset with compressed bond configurations creates more balanced training data and improves the

overall performance. More comprehensive comparisons of the pre-trained and fine-tuned ML prediction accuracy for Hessians is provided in Supplementary Figs. S4–S8. R: reactant; TS: transition state; P: product; RMSE: root mean squared error; DFT: density functional theory; ML: machine learning. Source data for this figure are provided with this paper.

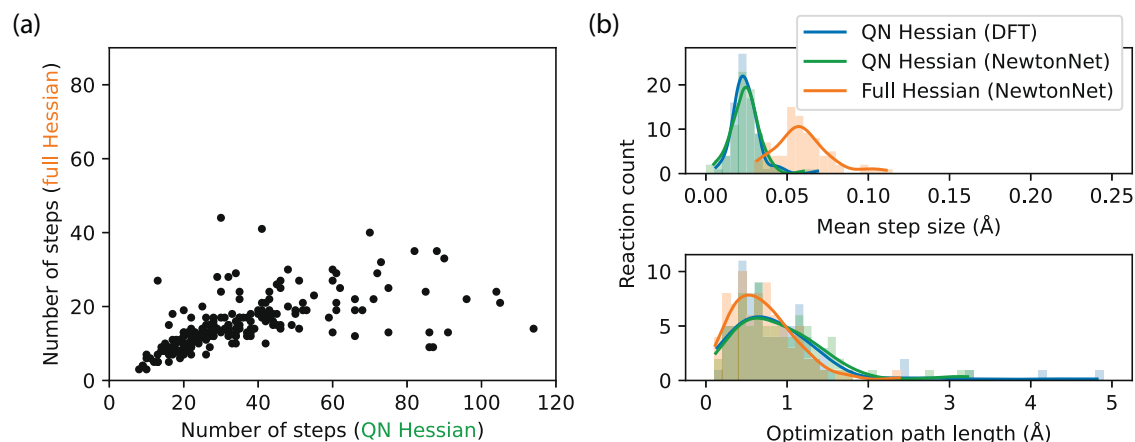


Fig. 3 | Efficiency improvement using full-Hessian TS optimization compared to the quasi-Newton approach. **a** The full-Hessian TS optimization requires 50% fewer steps to reach convergence than the QN approximate-Hessian approach, using the identical NewtonNet potential on the same reactions. **b** The improved efficiency of the full-Hessian TS optimization comes from both more confident steps (top) and more direct paths (bottom) to converge. In this efficiency

comparison, gradient calls for initial Hessian construction or Hessian reconstruction for QN restarts have been excluded, whether using DFT or NewtonNet for gradient calculations. TS: transition state; QN: quasi-Newton; DFT: density functional theory. Lines correspond to kernel density estimate fits to the histogram data. Source data for this figure are provided with this paper.

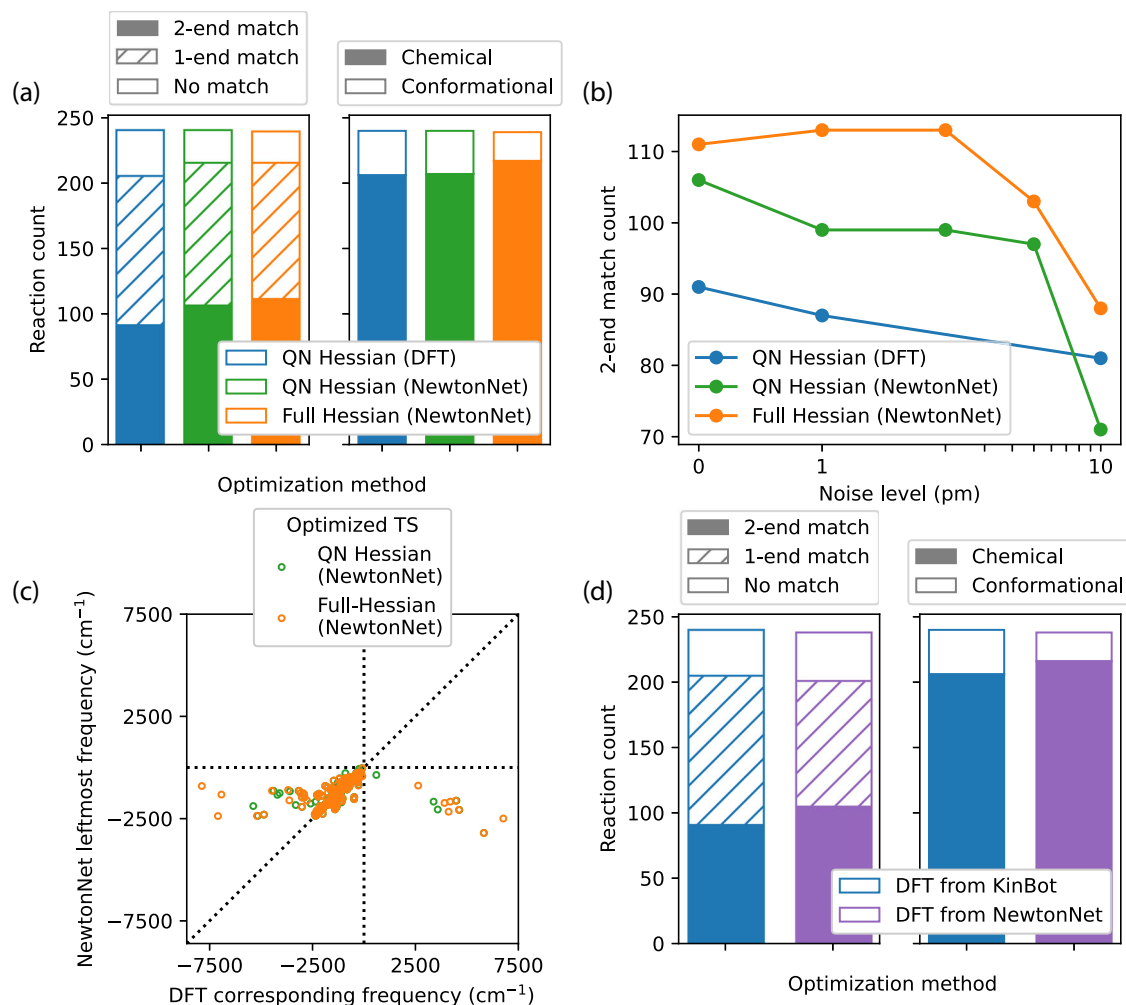


Fig. 4 | The quality of optimized TSs using NewtonNet. **a** NewtonNet predicts reactions that match the intended reactions with higher success rates, both with and without full Hessians, compared to DFT. The full Hessian also finds ~10% more TSs which involve chemical reactions as opposed to conformational changes. **b** The value of the full Hessians over the approximate QN approach is apparent when the quality of the initial TS guesses deteriorates. The QN convergence decays with additional noise to the guess structures, while the full Hessian convergence is more

robust to perturbations. **c** Comparing whether the left most frequency found on the ML PES is also a negative frequency on the DFT PES using the ML geometry. **d** Reoptimizing the ML transition state structure on the DFT surface demonstrates superior performance for 2-end matches and identifying chemical reactions compared with starting from the original KinBot initial guess. TS: transition state; QN: quasi-Newton; DFT: density functional theory; ML: machine learning; PES: potential energy surface. Source data for this figure are provided with this paper.

quantitatively illustrates the advantage of using full Hessians over QN Hessians but also shows that their performances can sometimes be qualitatively different. The QN approximation can make the optimization process unnecessarily difficult, even when the underlying problem is not inherently more complex. This observation strongly suggests that the poor convergence in TS optimization is more likely due to the Hessian approximation rather than the quality of the initial guesses or the complexity of the PESs.

Greater optimization efficiency can stem from two factors: increased confidence at each optimization step to increase the step size and a more optimal overall path of optimization. As shown in Fig. 3b, both factors improve the efficiency of the TS optimization with full analytical Hessians. In particular use of the full analytical Hessians exhibits an increase in confidence as measured by the RS-PRFO method, which allows for increased step size on average. A shorter, closer to optimal optimization path also plays a smaller but significant role with the analytical Hessians compared to the QN approach whether on the ML or DFT PES.

Next we consider the robustness of TS optimization using the fully analytical Hessians versus approximate Hessians (using ML or DFT) by

comparing the intended reactions from KinBot to those predicted from the IRC after TS optimization. As shown in Fig. 4a, the NewtonNet full or QN Hessian yielded the intended IRC reactant and product endpoints (2-end match) more often than the QN Hessian from DFT. In some cases only a 1-end match is found because the predicted product is more stable, and more chemically plausible, than the intended product from KinBot on a neutral and closed-shell PES. We also characterize the types of TSs found in Fig. 4a, in which the graphs of the reactants and products are inequivalent for chemical reactions, whereas isomorphic endpoints indicate a conformational TS; in either case, both converge to a first-order saddle point. All KinBot initial guess geometries are intended to yield chemical reaction TSs, and we see that employing full Hessians in TS optimization yields ~10% more TSs that involve chemical reactions as opposed to conformational changes compared to the approximate QN method.

The unreliability of a generated initial guess structure can lead to poor convergence or inaccurate predictions. Therefore, we consider a measure of robustness in which the TS optimization must recover from a poor starting structure, which we analyze by systematically introducing random structural perturbations to the guess structures

generated from KinBot. Figure 4b demonstrates the robustness of NewtonNet's full-Hessian optimization, maintaining consistent performance and even exhibiting slight improvements as noise levels increase up until 2–3 pm. In contrast, the performance using approximated QN Hessians rapidly decays, even using DFT, highlighting the importance of accurate Hessians for robust TS optimization.

A final important metric of robustness is whether NewtonNet predicted TS structures have negative eigenvalues on the DFT PES and/or improve the outcome of TS optimization on the DFT PES. In Fig. 4c we compare the vibrational frequencies of the NewtonNet optimized TS saddle point structures, using both the full Hessian and the QN Hessian, and use those structures as input to calculate the frequencies from the DFT Hessian. We then identify the DFT frequency mode that corresponds to the negative frequency mode from NewtonNet. For ~96% of the reactions, NewtonNet predicts highly accurate imaginary frequencies, regardless of full or QN Hessians. However, there are 10 cases for the QN-ML saddle points, and 11 for the full-ML Hessian saddle points, which have positive DFT Hessian eigenvalues. We further reoptimized all 240 reactions using the full Hessian ML TS structures as initial guesses on the DFT surface (Fig. 4d). In comparison with the optimization outcomes when starting with the original KinBot guess structures, we see an increase in both 2-end matches and chemical transformations under the DFT reoptimization. Thus, starting from ML optimized TS structures we find overall improved solutions on the DFT surface.

Discussion

We have presented a highly generalizable approach for predicting ab initio Hessians using machine learning based solely on energy and gradient data, and only requiring the property of second-order differentiability. Although Denzel and co-workers found that the feasibility of predicting Hessians using ML without access to explicit Hessian training data was poor³⁶, our study shows that solely relying on energy and force data using a well-trained ML model can efficiently and accurately predict the Hessian for reactive systems. We attribute our contrasting conclusions to the sufficiently high-quality PES obtained through the deep equivariant message-passing neural network and the mathematical relationship between the potential energy and its derivatives using a ML model that is C^2 continuous. To our knowledge, none of the widely used chemical datasets currently include Hessian information, and it is unlikely that such datasets will become available in the foreseeable future due to the high cost of generating ab initio Hessians. Thus, it is good news that models trained on energies and forces are sufficient to derive meaningful ML Hessians.

The ability to generate high-quality explicit Hessians with deep learning models obviates the complexities and assumptions associated with standard TS optimization approaches, in which 240 new reactions never seen in the training data are predicted with greater efficiency, accuracy, and robustness compared to QN ML or DFT. The implementation in Sella to utilize the ML Hessians incurs minimal additional computational overhead and requires no model retraining. Using NewtonNet, the Hessians can be calculated at least three orders of magnitude faster than the DFT calculation, while use of the full ML Hessian takes 2–3 \times fewer steps in the TS optimization compared to the QN approach.

This work emphasizes a TS state search methodology and hence used available training data from ANI-1, ANI-1x, and T1x that is specialized for applications involving reactive molecular organic systems but at a low level of DFT and basis set quality. Hence, for accurate predictions it would be desirable to recalculate these data sets at a higher level of theory, likely better density functionals and certainly larger basis sets, in order to predict quantitative barriers. Of course, with appropriate new data sets, we can generalize the ML-Hessian

approach for practical applications of TS optimization in many areas of chemical and material sciences. This is made possible by the tight software integration we have developed for NewtonNet with Sella, and workflows which could be trivially generalized to other relevant ML potentials.

We envision several areas of TS optimization using the ML approach we have described here. For example, the TS discovery for a reaction using methods such as NEB^{26,27}, QST²⁸, and GSM²⁹ often turns to local TS optimization methods when a reasonable approximate structure has been obtained¹². We showed that our ML transition state structures do improve the DFT optimizations. Therefore, a stepwise procedure that integrates mathematical optimization and machine learning to collaboratively achieve TS discovery for chemical reactions should be viable. In addition, since we can accurately calculate vibrational frequencies at the optimized TS structures, molecular free energy can be efficiently derived using the harmonic approximation. Further, since the quadratic correction can be applied at each step throughout the optimization process, it is possible to optimize TS structures on the free energy surface rather than PES, extending the feasibility of variational transition state theory (VTST) to larger systems where ab initio vibrational analysis becomes impractical³¹.

Methods

Data preparation

The T1x dataset for training is split in two different ways to assess the accuracy of the model predictions, illustrated in Supplementary Figs. S2 and S3. The original splitting in the literature is based on molecular compositions. All geometries with the same formula (equilibrium and non-equilibrium) are part of the same training, validation, and test set, leading to minimum data leakage. The error from this splitting can be regarded as the worst-case uncertainty estimation of an unseen configuration for the application on the real system in Figs. 1 and 2. On the other hand, we wish to maximize the chemical knowledge from the dataset learned by our model, so a more conventional splitting among molecular conformations is devised to learn the PES of all reactions. Hence in the design of the test set, we ensured that no reaction had both reactant and product pairs found in the training set.

The reference DFT Hessians are performed using Q-Chem 6.0.0⁵², using the ω B97X functional⁴¹ and 6–31G* basis set⁵³ in order to maintain compatibility with the T1x dataset³⁷. The eigenvalues are assigned based on the cosine similarity between the predicted and reference eigenvectors as a linear sum assignment problem using the Jonker–Volgenant algorithm⁵⁴.

The performance of the ML-Hessian and ML- and DFT-QN TS optimizations is evaluated by the Sella benchmark dataset³⁹. The dataset contains 500 small organic molecules between 7 and 25 atoms in configurations that approximate TS geometries across reaction families, among which 265 reactions are closed shell. However, 25 are present in the T1x dataset, thus we subsequently exclude those. We regenerate the remaining 240 such that, in contrast to the original data, the guess structures in our work are constrained minima on an ab initio PES instead of saddles on a semi-empirical PES³⁷. We also inject Gaussian noise up to 50 pm directly onto the atomic positions in the Cartesian coordinates of the initial guesses in order to degrade them for the purpose of understanding a given methods ability to still find the TS.

NewtonNet model and training details

The NewtonNet model with 3 message passing layers is trained using the same architecture as described in Reference³². Each node encodes an atomic environment into 128 features initialized by atom types Z_i , and each edge encodes an interatomic distance R_{ij} in 20 radial basis functions with a polynomial cutoff of 5 Å^{45,55}. The node features are

equivariantly updated with messages from neighboring nodes and edges. The molecular energy \tilde{E} is the sum of all atomic energies \tilde{E}_i ⁵⁶,

$$\tilde{E} = \sum_i^A \tilde{E}_i(\{Z_i\}, \{R_{ij}\}) \quad (1)$$

where atomic energies \tilde{E}_i are predicted from the node features at the final layer, and A is the total number of atoms. The predicted atomic forces \tilde{F}_i are calculated as the first derivative of the molecular energy \tilde{E} with respect to atomic positions R_i ⁵⁷,

$$\tilde{F}_i = -\nabla_i \tilde{E} = -\frac{\partial \tilde{E}}{\partial R_i} \quad (2)$$

and the predicted atomic Hessians \tilde{H}_{ij} are further calculated as the second analytical derivatives of the energy,

$$\tilde{H}_{ij} = \frac{\partial^2 \tilde{E}}{\partial R_j \partial R_i} = \frac{\partial^2 \tilde{E}}{\partial R_i \partial R_j} \quad (3)$$

However, due to the lack of training data for Hessians H_{ij} , only the energy \tilde{E} and forces \tilde{F}_i are trained in the loss function \mathcal{L} ,

$$\mathcal{L} = \frac{\lambda_E}{M} \sum_m^M (\tilde{E}_m - E_m)^2 + \frac{\lambda_F}{M} \sum_m^M \frac{1}{3A_m} \sum_i^{A_m} \|\tilde{F}_{mi} - F_{mi}\|^2 \quad (4)$$

where M is the total number of molecular graphs, which is 8 million for training, 1 million for validation, and 1 million for testing. After training on energy prediction \tilde{E} and force prediction \tilde{F}_i , the model is applied to infer Hessians H_{ij} without further training or fine-tuning.

We use a mini-batch gradient descent algorithm with a batch size of 100 to minimize the loss function using the Adam optimizer⁵⁸ with an initial learning rate of 10^{-4} and a decay rate of 0.7 on plateau. Fully connected neural networks with sigmoid linear unit (SiLU) nonlinearity⁴⁴ for all functions were used throughout the message passing layer. The application of smooth activation functions like SiLU is critical because the network has to be at least twice differentiable for Hessian calculations. We take $\lambda_E = 1$ and $\lambda_F = 20$ in the loss function in Equation (4) to put more emphasis on forces for derivative properties, and an additional L2 regularization of 10^{-5} is applied on all trainable parameters to further smooth out the potential energy surface. Layer normalization⁵⁹ on the atomic features at every message passing layer is applied for the stability of training. An ensemble of four models is trained on each splitting manner to ensure the reproducibility and reliability of the prediction. An outlier among the 4 predictions is removed if its absolute difference from the closest number compared to the difference from farthest number is larger than the 95% confidence limit of the Dixon Q's test⁴⁰.

Transition state optimization

For the transition state calculations, we use the Quantum Accelerator (QuAcc)⁶⁰, a Python package for high-throughput quantum chemistry workflows with an easy-to-use interface for Atomic Simulation Environment (ASE)⁶¹ optimizers. We utilize Sella³⁹ as the ASE optimizer for TS and intrinsic reaction coordinate (IRC) calculations, having implemented the feature to provide an external Hessian matrix at each optimization step.

Using Sella, the Hessian is automatically transformed into internal coordinates and iteratively diagonalized using the Rayleigh–Ritz procedure for the leftmost eigenpair by a modified Jacobi–Davidson method (JDO, or Olsen's method), with finite difference step size of 10^{-4} Å and convergence threshold of 0.1¹². The TS optimization steps are determined by restricted step partitioned rational function optimization (RS-PRFO)^{48–50}. The trust radius is initially 0.1 and adjusted based on the improper ratio (>1) between the predicted and actual

energy change; the radius is increased by a factor of 1.15 when the ratio is below 1.035 and decreased by a factor of 0.65 when the ratio is above 5.0. The IRC is determined by energy minimization at a trust radius of 0.1 Å/amu^{-1/2} in mass-weighted coordinates⁶². QN Hessian updates are achieved using TS-BFGS^{22,24}. A maximum of 1000 steps is applied for both TS optimization and IRC search.

The comparison of reactants and products is based on graph isomorphism. Molecular connectivity graphs are created using Open Babel⁶³ with atom indexing and compared using the VF2 algorithm⁶⁴. The optimization path length is calculated in the Cartesian coordinate distance with the Kabsch algorithm⁶⁵. The path length in Fig. 3b only accounts for reactions with 2-end matches.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data⁶⁶ including initial transition state guess structures, optimized transition states, and corresponding reactants and products with their coordinates of geometry, energy, forces and Hessians are available at <https://doi.org/10.6084/m9.figshare.25356616>. Source data for Figs. 1–4 is available with this manuscript. Source data are provided in this paper.

Code availability

The codebase is comprised of several publicly available packages and tools that contribute to the project. Sella³⁹ is publicly accessible at <https://github.com/zadorlab/sella> and comes with comprehensive documentation. NewtonNet⁶⁷, another integral part of the project, is also publicly available at <https://github.com/THGLab/NewtonNet>. The recipes implemented in QuAcc⁶⁰ for NewtonNet and Q-Chem, utilizing Sella as the ASE optimizer for transition state and IRC calculations, are publicly accessible and accompanied by thorough documentation at <https://github.com/Quantum-Accelerators/quacc>. The full workflow and the analysis scripts⁶⁸, responsible for generating molecular graphs, retrieving data from the MongoDB database, and performing graph isomorphisms to analyze reactions, are available at <https://github.com/THGLab/MLHessian-TSopt>. This comprehensive summary provides insights into the availability of the codebase for potential readers and collaborators. Examples to use our end-to-end workflow⁶⁹ are available at <https://github.com/kumarantu/ts-workflow-examples>.

References

- Davidson, E. R. The iterative calculation of a few of the lowest eigenvalues and corresponding eigenvectors of large real-symmetric matrices. *J. Comp. Phys.* **17**, 87–94 (1975).
- Amos, R. D. & Rice, J. E. Implementation of analytic derivative methods in quantum chemistry. *Comp. Phys. Rep.* **10**, 147–187 (1989).
- Barter, D. et al. Predictive stochastic analysis of massive filter-based electrochemical reaction networks. *Digital Discovery* **2**, 123–137 (2023).
- Spotte-Smith, E. W. C. et al. Chemical Reaction Networks Explain Gas Evolution Mechanisms in Mg-Ion Batteries. *J. Am. Chem. Soc.* **145**, 12181–12192 (2023).
- Nelder, J. A. & Mead, R. A simplex method for function minimization. *Comput. J.* **7**, 308–313 (1965).
- Brent, R. *An Algorithm with Guaranteed Convergence for Finding a Zero of a Function*. (Prentice-Hall, Englewood Cliffs, NJ, 1973).
- Debye, P. Näherungsformeln für die zylinderfunktionen für große werte des arguments und unbeschränkt veränderliche werte des index. *Math. Ann.* **67**, 535–558 (1909).
- Hestenes, M. R. & Stiefel, E. Methods of conjugate gradients for solving linear systems. *J. Res. Natl Bur. Stand.* **49**, 409–435 (1952).

9. Nocedal, J. & Wright, S. J. *Numerical Optimization*. 2nd edn. (Springer, New York, NY, USA, 2006).
10. Baker, J. An algorithm for the location of transition states. *J. Comp. Chem.* **7**, 385–395 (1986).
11. Schlegel, H. B. Geometry optimization. *Wires.: Comput. Mol. Sci.* **1**, 790–809 (2011).
12. Hermes, E. D., Sargsyan, K., Najm, H. N. & Zádor, J. Accelerated saddle point refinement through full exploitation of partial hessian diagonalization. *J. Chem. Theo. Comput.* **15**, 6536–6549 (2019).
13. Császár, P. & Pulay, P. Geometry optimization by direct inversion in the iterative subspace. *J. Mol. Struct.* **114**, 31–34 (1984).
14. Schlegel, H. B. Optimization of equilibrium geometries and transition structures. *J. Comp. Chem.* **3**, 214–218 (1982).
15. Schlegel, H. B. Estimating the hessian for gradient-type geometry optimizations. *Theo. Chim. Acta* **66**, 333–340 (1984).
16. Schlegel, H. B. *Optimization of Equilibrium Geometries and Transition Structures*, 249–286 (John Wiley & Sons, Ltd, 1987).
17. Fischer, T. H. & Almlöf, J. General methods for geometry and wave function optimization. *J. Phys. Chem.* **96**, 9768–9774 (1992).
18. Lindh, R., Bernhardsson, A., Karlström, G. & Malmqvist, P.-A. On the use of a hessian model function in molecular geometry optimizations. *Chem. Phys. Lett.* **241**, 423–428 (1995).
19. Jensen, F. Using force fields methods for locating transition structures. *J. Chem. Phys.* **119**, 8804–8808 (2003).
20. Chantreau Majerus, R., Robertson, C. & Habershon, S. Assessing and rationalizing the performance of hessian update schemes for reaction path hamiltonian rate calculations. *J. Chem. Phys.* **155**, 204112 (2021).
21. Dennis Jr, J. E. & Schnabel, R. B. *Numerical Methods For Unconstrained Optimization And Nonlinear Equations*. (SIAM, 1996).
22. Anglada, J. M. & Bofill, J. M. How good is a broyden-fletcher-goldfarb-shanno-like update hessian formula to locate transition structures? specific reformulation of broyden-fletcher-goldfarb-shanno for optimizing saddle points. *J. Comp. Chem.* **19**, 349–362 (1998).
23. Fletcher, R. *Practical methods of optimization* (John Wiley and Sons, 2000).
24. Bofill, J. M. Remarks on the updated hessian matrix methods. *Int. J. Quant. Chem.* **94**, 324–332 (2003).
25. Hratchian, H. P. & Schlegel, H. B. Using hessian updating to increase the efficiency of a hessian based predictor-corrector reaction path following method. *J. Chem. Theo. Comput.* **1**, 61–9 (2005).
26. Jonsson, H., Mills, G. & Jacobsen, K. W. *Nudged Elastic Band Method For Finding Minimum Energy Paths Of Transitions*, 385–404 (World Scientific, 1998).
27. Henkelman, G., Uberuaga, B. P. & Jónsson, H. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J. Chem. Phys.* **113**, 9901–9904 (2000).
28. Govind, N., Petersen, M., Fitzgerald, G., King-Smith, D. & Andzelm, J. A generalized synchronous transit method for transition state location. *Computational Mater. Sci.* **28**, 250–258 (2003).
29. Peters, B., Heyden, A., Bell, A. T. & Chakraborty, A. A growing string method for determining transition states: comparison to the nudged elastic band and string methods. *J. Chem. Phys.* **120**, 7877–7886 (2004).
30. Zhang, Y. et al. Dp-gen: a concurrent learning platform for the generation of reliable deep learning based potential energy models. *Comp. Phys. Comm.* **253**, 107206 (2020).
31. Bac, S., Patra, A., Kron, K. J. & Mallikarjun Sharada, S. Recent advances toward efficient calculation of higher nuclear derivatives in quantum chemistry. *J. Phys. Chem. A* **126**, 7795–7805 (2022).
32. Haghghatdari, M. et al. Newtonnet: a newtonian message passing network for deep learning of interatomic potentials and forces. *Dig. Disc.* **1**, 333–343 (2022).
33. Duan, C., Du, Y., Jia, H. & Kulik, H. J. Accurate transition state generation with an object-aware equivariant elementary reaction diffusion model. *Nat. Comp. Sci.* **3**, 1045–1055 (2023).
34. Kim, S., Woo, J. & Kim, W. Y. Diffusion-based generative ai for exploring transition states from 2d molecular graphs. *Nat. Comm.* **15**, 341 (2024).
35. Denzel, A. & Kästner, J. Gaussian process regression for transition state search. *J. Chem. Theo. Comput.* **14**, 5777–5786 (2018).
36. Denzel, A. & Kästner, J. Hessian matrix update scheme for transition state search based on gaussian process regression. *J. Chem. Theo. Comput.* **16**, 5083–5089 (2020).
37. Schreiner, M., Bhowmik, A., Vegge, T., Busk, J. & Winther, O. Transition1x - a dataset for building generalizable reactive machine learning potentials. *Sci. Data* **9**, 779 (2022).
38. Van de Vijver, R. & Zádor, J. Kinbot: automated stationary point search on potential energy surfaces. *Comp. Phys. Comm.* **248**, 106947 (2020).
39. Hermes, E. D., Sargsyan, K., Najm, H. N. & Zádor, J. Sella, an open-source automation-friendly molecular saddle point optimizer. *J. Chem. Theo. Comput.* **18**, 6974–6988 (2022).
40. Guan, X., Heindel, J. P., Ko, T., Yang, C. & Head-Gordon, T. Using machine learning to go beyond potential energy surface benchmarking for chemical reactivity. *Nat. Comp. Sci.* **3**, 965–974 (2023).
41. Chai, J.-D. & Head-Gordon, M. Systematic optimization of long-range corrected hybrid density functionals. *J. Chem. Phys.* **128**, 084106 (2008).
42. Smith, J. S., Isayev, O. & Roitberg, A. E. Ani-1: an extensible neural network potential with dft accuracy at force field computational cost. *Chem. Sci.* **8**, 3192–3203 (2017).
43. 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 (2018).
44. Elfving, S., Uchibe, E. & Doya, K. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. *Neural Netw.* **107**, 3–11 (2018).
45. Gasteiger, J., Groß, J. & Günnemann, S. Directional message passing for molecular graphs. In *ICLR 2020*. <http://arxiv.org/abs/2003.03123> (2022).
46. Sharada, S. M., Bell, A. T. & Head-Gordon, M. A finite difference davidson procedure to sidestep full ab initio hessian calculation: Application to characterization of stationary points and transition state searches. *J. Chem. Phys.* **140**, 164115 (2014).
47. Hermes, E. D., Sargsyan, K., Najm, H. N. & Zádor, J. Geometry optimization speedup through a geodesic approach to internal coordinates. *J. Chem. Phys.* **155**, 094105 (2021).
48. Banerjee, A., Adams, N., Simons, J. & Shepard, R. Search for stationary points on surfaces. *J. Phys. Chem.* **89**, 52–57 (1985).
49. Anglada, J. M. & Bofill, J. M. A reduced-restricted-quasi-newton-raphson method for locating and optimizing energy crossing points between two potential energy surfaces. *J. Comp. Chem.* **18**, 992–1003 (1997).
50. Besalú, E. & Bofill, J. M. On the automatic restricted-step rational-function-optimization method. *Theo. Chem. Acc.* **100**, 265–274 (1998).
51. Bao, J. L. & Truhlar, D. G. Variational transition state theory: theoretical framework and recent developments. *Chem. Soc. Rev.* **46**, 7548–7596 (2017).
52. Epifanovsky, E. et al. Software for the frontiers of quantum chemistry: an overview of developments in the q-chem 5 package. *J. Chem. Phys.* **155**, 084801 (2021).
53. Ditchfield, R., Hehre, W. J. & Pople, J. A. Self-consistent molecular-orbital methods. ix. an extended gaussian-type basis for molecular-orbital studies of organic molecules. *J. Chem. Phys.* **54**, 724–728 (2003).

54. Crouse, D. F. On implementing 2d rectangular assignment algorithms. *IEEE Trans. Aerosp. Electron. Syst.* **52**, 1679–1696 (2016).
55. Schütt, K. T., Unke, O. T. & Gastegger, M. Equivariant message passing for the prediction of tensorial properties and molecular spectra. *Proc. 38th Int. Conf. Mach. Learn.* **139**, 9377–9388 (2021).
56. Behler, J. & Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98**, 146401 (2007).
57. Chmiela, S., Sauceda, H. E., Müller, K.-R. & Tkatchenko, A. Towards exact molecular dynamics simulations with machine-learned force fields. *Nat. Comm.* **9**, 3887 (2018).
58. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. In *the 3rd International Conference for Learning Representations*. <http://arxiv.org/abs/1412.6980> (2017).
59. Ba, J. L., Kiros, J. R. & Hinton, G. E. Layer normalization. In *NIPS 2016 Deep Learning Symposium*. <http://arxiv.org/abs/1607.06450> (2016).
60. Rosen, A. quacc – the quantum accelerator. <https://zenodo.org/records/13139853> (2024).
61. Larsen, A. H. et al. The atomic simulation environment—a python library for working with atoms. *J. Phys.: Cond. Matt.* **29**, 273002 (2017).
62. Müller, K. & Brown, L. D. Location of saddle points and minimum energy paths by a constrained simplex optimization procedure. *Theoretica Chim. acta* **53**, 75–93 (1979).
63. O’Boyle, N. M. et al. Open babel: an open chemical toolbox. *J. Cheminform.* **3**, 33 (2011).
64. Foggia, P., Sansone, C. & Vento, M. An improved algorithm for matching large graphs. In *3rd IAPR-TC15 workshop on graph-based representations in pattern recognition*. (2001).
65. Kabsch, W. A solution for the best rotation to relate two sets of vectors. *Acta Cryst. Sec. A* **32**, 922–923 (1976).
66. Yuan, E. et al. Data for deep learning of ab initio Hessians for transition state optimization. <https://doi.org/10.6084/m9.figshare.25356616.v1> (2024).
67. Yuan, E., Haghighatlari, M., Rosen, A. S., Guan, N. X. & Jerry-JohnsonLee. Thglab/newtonnet: v1.0.1. <https://zenodo.org/records/13130421> (2024).
68. Yuan, E. et al. Thglab/mlhessian-tsopt: v1.0.2. <https://zenodo.org/records/13128544> (2024).
69. Kumar, A. et al. ericyuan00000/ts-workflow-examples: v1.0.0. <https://zenodo.org/records/13128509> (2024).

Acknowledgements

S.M.B., A.K., E.C.-Y.Y., and T.H.-G. thank the Lawrence Berkeley National Laboratory Director Research Development program for the work supporting transition state methods. E.C.-Y.Y., X.G., and T.H.-G. thank the CPIMS program, Office of Science, Office of Basic Energy Sciences, and Chemical Sciences Division of the U.S. Department of Energy under Contract DE-AC02-05CH11231 for support of the machine learning. E.D.H. and J.Z. acknowledge the Exascale Catalytic Chemistry (ECC) Project supported by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, Chemical Sciences, Geosciences, and Biosciences Division, as part of the Computational Chemistry Sciences Program for work related to Sella. J.Z. acknowledges the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Division of Chemical Sciences, Geosciences, and Biosciences under the

Gas Phase Chemical Physics program for work related to KinBot. A.S.R. acknowledges support via a Miller Research Fellowship from the Miller Institute for Basic Research in Science, University of California, Berkeley. This work used computational resources provided by the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility operated under Contract DE-AC02-05CH11231, and the Lawrence computational cluster resource provided by the IT Division at the Lawrence Berkeley National Laboratory (Supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231).

Author contributions

S.M.B. and T.H.G. designed the project. E.C.-Y.Y. and X.G. carried out the NewtonNet training and Hessian calculations, E.D.H., A.K. and S.M.B. interfaced NewtonNet with the Sella software package, A.K., S.M.B., and A.S.R. implemented workflows, A.K. and S.M.B. executed TS workflows, J.Z. ran KinBot to generate initial guess structures and reference end-points, and E.C.-Y.Y. and T.H.G. wrote the paper. All authors discussed the results and made comments and edits to the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-52481-5>.

Correspondence and requests for materials should be addressed to Teresa Head-Gordon or Samuel M. Blau.

Peer review information *Nature Communications* thanks Chenru Duan and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024