

# Accelerating Molecular Dynamics Simulations with Foundation Neural Network Models Using Multiple Time Steps and Distillation

Côme Cattin, Thomas Plé,\* Olivier Adjoua, Nicolaï Gouraud, Louis Lagardère, and Jean-Philip Piquemal\*



Cite This: <https://doi.org/10.1021/acs.jpcllett.5c03720>



Read Online

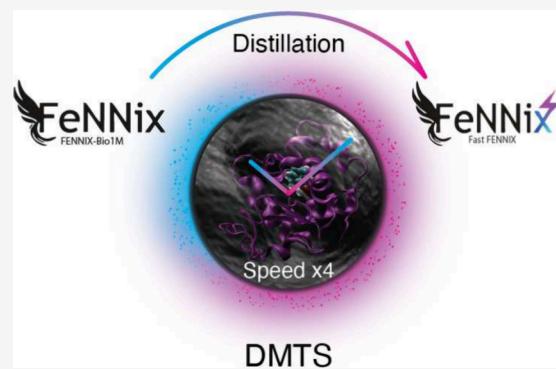
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

**ABSTRACT:** We present a distilled multi-time-step (DMTS) strategy to accelerate molecular dynamics simulations using foundation neural network models. DMTS uses a dual-level neural network, where the target accurate potential is coupled to a simpler but faster model obtained via a distillation process. The 3.5 Å cutoff distilled model is sufficient to capture the fast-varying forces, i.e., mainly bonded interactions, from the accurate potential, allowing its use in a reversible reference system propagator algorithm (RESPA)-like formalism. The approach conserves accuracy, preserving both static and dynamic properties, while enabling us to evaluate the costly model only every 3 to 6 fs depending on the system. Consequently, large simulation speedups over standard 1 fs integration are observed: nearly 4-fold in homogeneous systems and 3-fold in large solvated proteins through leveraging active learning for enhanced stability. Such a strategy is applicable to any neural network potential and reduces the performance gap with classical force fields.



**N**eural Network potentials (NNPs)<sup>1–15</sup> have emerged as powerful tools to perform molecular dynamics (MD) simulations, offering near-quantum mechanical accuracy at a fraction of the computational cost of *ab initio* methods. Presently, such models can be generalized up to Foundation Models covering complete area of applications of molecular dynamics ranging from Material Science to Chemistry, Biology, and Drug Design.<sup>16–20</sup> By learning the potential energy surface from reference quantum chemistry data, these models enable simulations of large and complex molecular systems with significantly improved fidelity compared to classical force fields.<sup>21</sup> However, this increase in accuracy comes at a cost: NNPs are substantially more expensive to evaluate than traditional empirical potentials. This limits their applicability in terms of system sizes and long-time-scale simulations.

One of the primary bottlenecks of performing MD with ML potentials lies in the time integration itself. The equations of motion must be solved with a small time step, typically on the order of 0.5 to 1 fs, to resolve high-frequency motions such as bond vibrations. With computationally intensive ML models, this results in a large number of expensive force evaluations, further amplifying the overall simulation cost.

Multi-Time-Step (MTS) integrators<sup>22–24</sup> offer a well-established strategy to address this challenge. Originally introduced in the context of classical molecular simulations, MTS methods such as RESPA (Reference System Propagator Algorithm)<sup>22</sup> exploit the separation of time scales between different components of the forces to reduce the number of expensive evaluations. By integration of fast-changing forces

with a small time step and slow-changing ones with a larger time step, significant performance gains can be achieved without compromising accuracy. Having proven their usefulness for classical force fields simulations, MTS schemes have been shown to be also applicable to *Ab Initio* MD (AIMD).<sup>25</sup> In the context of machine learning molecular simulations, few works were reported in that direction.<sup>26–29</sup> Thus, MTS schemes are not presently part of today's NNP toolkit. Indeed, in their case, force decomposition is not naturally given by physical interactions and naive MTS implementations can be plagued by various sources of instabilities, for example resonances.<sup>30,31</sup> Still, MTS represents an important opportunity for foundation models, as the use of multiple neural networks of differing complexity and inference cost could enable efficient MTS schemes specifically tailored to ML potentials.

In this work, we propose the first global MTS strategy for neural network foundation models. As it relies on knowledge distillation,<sup>32</sup> we term it DMTS for distilled multi-time-step. It is applicable to any type of system and ensures stable long time scale simulations with consequent speed gains. We leverage the

Received: November 26, 2025

Revised: January 7, 2026

Accepted: January 9, 2026

## Chart 1

**Algorithm 1** MTS Integration Step with FENNIX Force Splitting

---

```

1: if first_step then
2:    $F_{\text{small}} \leftarrow \text{FENNIX}_{\text{small}}(x)$ 
3:    $F \leftarrow \text{FENNIX}_{\text{large}}(x)$ 
4: end if
5:  $v \leftarrow v + \frac{\Delta t}{2m} \cdot (F - F_{\text{small}})$ 
6: for  $i = 1$  to  $n_{\text{slow}}$  do
7:    $v \leftarrow v + \frac{\Delta t}{2m \cdot n_{\text{slow}}} \cdot F_{\text{small}}$ 
8:    $x \leftarrow x + \frac{\Delta t}{2 \cdot n_{\text{slow}}} \cdot v$ 
9:    $v \leftarrow \text{thermo}(v, \frac{\Delta t}{n_{\text{slow}}})$                                 ▷ Apply thermostat
10:   $x \leftarrow x + \frac{\Delta t}{2 \cdot n_{\text{slow}}} \cdot v$ 
11:   $F_{\text{small}} \leftarrow \text{FENNIX}_{\text{small}}(x)$ 
12:   $v \leftarrow v + \frac{\Delta t}{2m \cdot n_{\text{slow}}} \cdot F_{\text{small}}$ 
13: end for
14:  $F \leftarrow \text{FENNIX}_{\text{large}}(x)$ 
15:  $v \leftarrow v + \frac{\Delta t}{2m} \cdot (F - F_{\text{small}})$ 

```

---

RESPA formalism since it preserves symplecticity and time-reversibility.<sup>22</sup> We integrate this approach into the FeNNol library,<sup>33</sup> a neural network framework for molecular simulations that can be coupled to the scalable Deep-HP machine learning interface<sup>26</sup> present in the GPU-accelerated Tinker-HP molecular dynamics package.<sup>34,35</sup> Our method associates a small, fast NNP obtained via distillation with the FeNNix-Bio1(M) foundation model,<sup>19</sup> leveraging the speed of the former to reduce the frequency of the expensive evaluations of the latter. We systematically evaluate the method using various systems including liquid water, solvated small molecules, and proteins in the condensed phase, exploring different hyperparameter settings and model sizes in order to design the fastest setup while maintaining accuracy. Throughout, we compare the relative merits of both system-specific and general-purpose models as a fast component for the DMTS scheme. We also discuss the use of active learning to systematically increase DMTS's accuracy and performance.

Our results show that this hybrid RESPA-like scheme enables substantial computational speed gains with limited loss of accuracy, offering a practical route to scalable and efficient Molecular Dynamics with foundation machine learning models and NNPs in general.

In this work, we use two neural network potentials of differing complexity to implement a MTS integration scheme. In this scheme, the dynamics of a cheaper NNP is integrated with a small time step ( $\sim 1$  fs) and is periodically corrected (every 2 to 6 steps depending on the system) by the force difference between the small NNP and a larger reference model, here the FeNNix-Bio1(M) model.<sup>19</sup> This enables us to recover the dynamics of the larger NNP without evaluating its forces at every time step, thus improving the computational efficiency. In particular, we employ the BAOAB-RESPA integration scheme<sup>24</sup> outlined in Algorithm 1 (Chart 1), that we implemented in the FeNNol library. For simulations in Tinker-HP, we use the implementation described in ref 24 and the Deep-HP interface<sup>26</sup> for calling models.

In Algorithm 1 (Chart 1),  $x$  denotes the system's coordinates,  $v$  its velocities,  $\Delta t$  the outer time step,  $m$  the mass, and  $n_{\text{slow}}$  the number of inner steps with time step

$\Delta t/n_{\text{slow}}$ .  $\text{FENNIX}_{\text{large}}(x)$  denotes the reference FeNNix-Bio1(M) machine-learned force field evaluated at configuration  $x$  and  $\text{FENNIX}_{\text{small}}(x)$  denotes the cheaper model. We describe the neural network architectures and the distillation strategies employed to train the  $\text{FENNIX}_{\text{small}}$  model in the next sections.

The reference model that we intend to accelerate corresponds to the FeNNix-Bio1(M) model, trained on a broad and diverse data set. This model is based on a range-separated equivariant transformer architecture, where close-range and long-range interactions are described with different spatial resolutions and dedicated attention heads. Its receptive field is 11 Å in total via two message-passing interactions. Details about the full architecture are provided in ref 19. The second, lighter weight model that is used in inner steps of the RESPA scheme uses the same base architecture but with reduced capacity and removes the long-range attention heads in order to enable faster inference and lower computational cost. Its receptive field is only 3.5 Å (one message-passing interaction), making it much more short-sighted and allowing it to focus mostly on fast-varying “bonded” forces. The model hyperparameters that we used in the following numerical experiments are provided in Supporting Information.

We derive the  $\text{FENNIX}_{\text{small}}$  model from the reference one via a knowledge distillation procedure<sup>32,36</sup> where the distilled model is trained on data labeled with the FeNNix-Bio1(M) model instead of DFT. This ensures that the forces used in the inner loop of the RESPA scheme are as close to the reference ones as possible, minimizing the correction necessary to recover the correct dynamics and thus the frequency of its application. In practice, the small model comes in two flavors: either (i) as an on-the-fly system-specific model or (ii) as a generic model. We briefly describe here these two distillation strategies. Specific details about the training parameters can be found in Supporting Information.

## ■ SYSTEM-SPECIFIC MODEL

For each system, a reference data set is generated by running a short MD simulation (less than a nanosecond) using the

reference model. For proteins and, more generally, for large systems, a fragmentation strategy similar to the one proposed in ref 15 was employed to reduce the computational burden while retaining local structural information. We then evaluated the energies and forces of the collected frames with the reference model. The on-the-fly system-specific model was then trained on this data set, resulting in a model about 10 times faster than FeNNix-Bio1(M) with our current setup.

## ■ GENERIC MODEL

In addition to system-specific distilled models, we also propose a generic fast model trained on a chemically diverse data set, enabling broader applicability and faster deployment in new systems. To construct this transferable potential, we generated a training set by evaluating FeNNix-Bio1(M) on a subset of conformations from the SPICE2 data set,<sup>37,38</sup> which contains a wide variety of small organic molecules and biologically relevant complexes. The generic model was trained on the obtained data set, resulting in a transferable model that captures general chemical patterns and can be reused across systems. This generic model can be used directly in the inner loop of our DMTS scheme or serve as an initialization point for further fine-tuning on a target system. This offers a compromise between generality and accuracy, especially when system-specific data are limited or a rapid deployment is desired.

The results obtained with the DMTS scheme for each of the following experiments were compared with a single time step (STS) integrator using the FeNNix-Bio1(M) force field with a time step of 1 fs. The STS scheme employs a BAOAB Langevin integrator.<sup>39,40</sup> All simulations used a friction coefficient of 1 ps<sup>-1</sup> for coupling to the Langevin thermostat.

## ■ BULK WATER

First, to investigate the robustness of our models and integrator, we performed a series of stability tests on a small water box containing 648 atoms. Simulations were carried out using the system-specific model and the generic model, with external time steps ranging from 2 to 9 fs. For each configuration, we monitored dynamic and thermodynamic observables, including diffusion coefficients, kinetic and potential energies, and temperature. Diffusion coefficients were computed using Tinker<sup>41</sup> with the Einstein formula. The stability of the trajectory was evaluated over the full course of the 2 ns simulation with qualitative indicators reported in Table 1. Additional tests were performed with hydrogen mass repartitioning (HMR),<sup>42</sup> enabling larger time steps.

At a short external time step (2 fs), both specific and generic models produced stable trajectories with diffusion coefficients and temperatures close to the reference STS values. At 3 fs, simulations remained stable, although the generic and system-specific models displayed a marked increase in both diffusion coefficient and temperature, suggesting the onset of integration artifacts.

The introduction of HMR enables stable simulations up to 6 fs, with dynamic and thermodynamic properties remaining within reasonable agreement between the STS and DMTS integrator for both models. However, at 7 fs and beyond, instabilities systematically occurred, reflected by divergent values in Table 1 or nonphysical increases in diffusion and temperature. These results highlight that HMR provides a substantial extension of the accessible time step, effectively

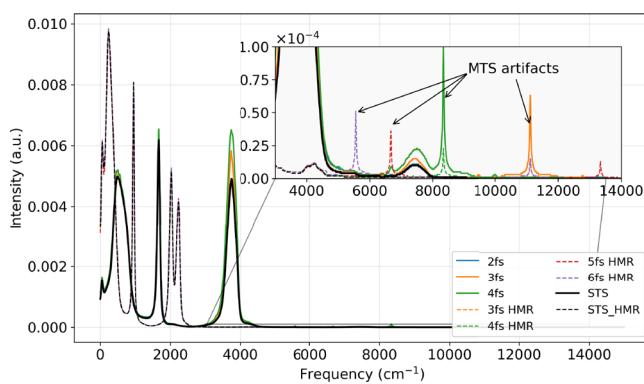
**Table 1. Stability Tests Performed on a Water Box of 648 Atoms Using Both the System-Specific and Generic Models at Various Integration Time Steps, With and without Hydrogen Mass Repartitioning (HMR)<sup>a</sup>**

	Diffusion		Temperature	
	System	Generic	System	Generic
STS	2.21 ± 0.15		300.1 ± 9.6	
2 fs	2.45 ± 0.17	2.06 ± 0.05	300.5 ± 9.6	300.2 ± 9.6
3 fs	2.30 ± 0.06	2.66 ± 0.67	315.2 ± 10.6	333.3 ± 12.2
3 fs HMR	2.27 ± 0.14	2.05 ± 0.18	300.1 ± 9.5	299.9 ± 9.5
4 fs HMR	2.11 ± 0.19	2.04 ± 0.24	300.8 ± 9.5	303.4 ± 9.9
5 fs HMR	2.30 ± 0.19	2.16 ± 0.25	302.7 ± 9.8	304.1 ± 10.2
6 fs HMR	2.20 ± 0.30	2.64 ± 0.47	304.5 ± 9.7	305.9 ± 9.9
7 fs HMR		3.32 ± 0.04		327.5 ± 11.1

<sup>a</sup>Diffusion coefficients (in  $1 \times 10^{-5}$  cm<sup>2</sup>/s) and average temperatures (K) are reported with associated standard deviations. Values are averaged over 2 ns trajectories. Instabilities prevented completion of simulations beyond 3 fs without HMR and beyond 7 fs with HMR.

doubling the stability range compared with standard masses. Nevertheless, care must be taken to avoid excessive increases, as instabilities rapidly emerge beyond 6 fs. In practice, time steps of 2–3 fs without HMR and up to 5–6 fs with HMR appear to offer the most reliable compromise between computational efficiency, stability, and physical accuracy.

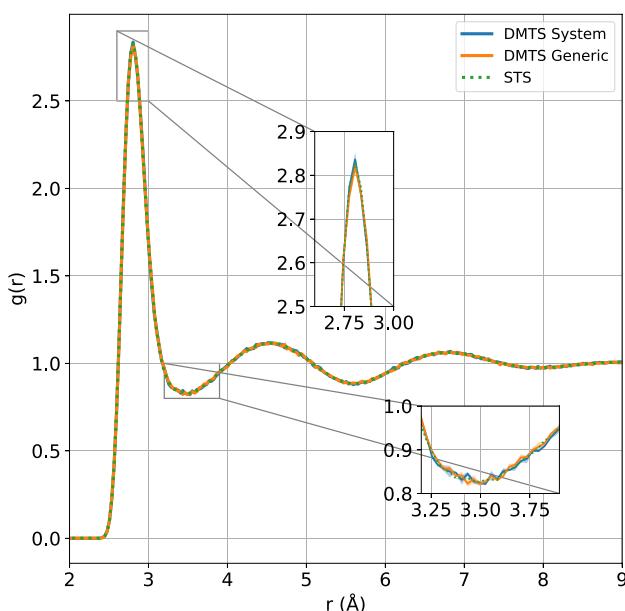
It is well documented that limits in the larger time step usable in the context of MTS schemes in molecular dynamics are related to the coupling of this largest time step with the highest-frequency vibrational modes involved in the system: the so-called resonance effects.<sup>30,31,43–48</sup> A simple diagnostic can be made to assess these aspects by considering the velocity autocorrelation spectra in this setup. Figure 1 illustrates this in the context of bulk water, where we see the artifacts produced by the periodic MTS force correction (annotated in the inset of Figure 1) progressively coupling with the overtone of the O–H stretching mode (around 7500 cm<sup>-1</sup> or 4000 cm<sup>-1</sup> with HMR) as the outer time step increases. This diagnostic



**Figure 1.** Average velocity autocorrelation spectra of hydrogen atoms in bulk water. The reference single time step (STS) simulation is shown in black. Distilled multi-time-step (DMTS) simulations are shown for different integration time steps (2, 3, 4, 5, and 6 fs), with corresponding heavy-mass repartitioning (HMR) variants represented by dashed lines of the same color. Inset shows zooms in high frequency regions containing MTS integration artifacts.

confirms the limits of the outer time step to 5 fs with hydrogen mass repartitioning and 2–3 fs without.

We further assess the robustness of our approach by computing the radial distribution function of a larger water box composed of 4800 atoms. As shown in Figure 2, the DMTS



**Figure 2.** Radial distribution function,  $g(r)$ , as a function of the distance ( $r$ ) in Å. The solid blue curve corresponds to the DMTS simulation with the on-the-fly system-specific model, and the orange one corresponds to DMTS with the small generic model. Both DMTS simulations used an inner time step of 1 fs and an outer time step of 5 fs. The dotted green curve corresponds to the reference STS simulation.

simulations (with both generic and system-specific models) with an outer time step of 5 fs and HMR correctly reproduce the STS results within the statistical uncertainties. Overall, with this setup, we obtained a speedup of around 4 compared to STS with a 1 fs time step for bulk water simulations corresponding to an increase from 6.59 to 25.03 ns/day (see Table 2).

## SOLVATED MOLECULES

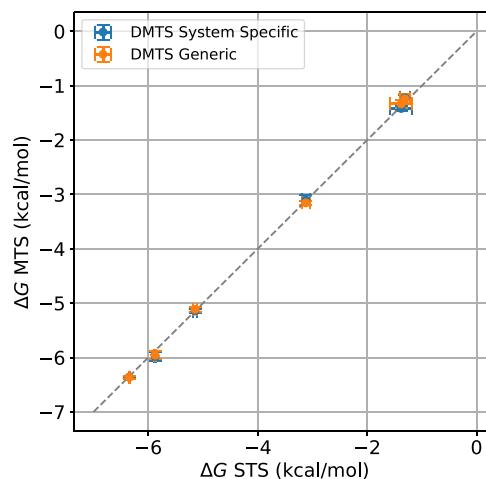
To further assess the accuracy of the proposed DMTS scheme, we turned to small solvated molecules. We observed stability

**Table 2. Maximum Performance (ns/day) on Small (648 atoms) and Large (4800 atoms) Water Boxes Obtained on a Single NVIDIA A100 GPU Using Both the System-Specific and Generic Models at Various Integration Time Steps, with and without Hydrogen Mass Repartitioning (HMR)**

	Small		Large	
	System	Generic	System	Generic
STS		38.07		6.59
2 fs	48.05	40.58	11.61	10.91
3 fs	67.81	57.57	16.26	15.26
4 fs HMR	85.92	71.73	20.57	19.22
5 fs HMR	98.69	84.85	25.03	22.81
6 fs HMR	117.99	95.67	27.52	26.03

limits similar to those for bulk water for the five molecules that we tested: ethanol, benzene, trimethylamine, diethylsulfide, and acetic acid. We then evaluated the hydration free energies (HFEs) of these (as well as those of water). Calculations were performed leveraging the alchemical Lambda-ABF<sup>49</sup> method, employing both system-specific models trained on-the-fly and a generic small foundation model. All simulations were carried out using the same DMTS integration scheme described above, with an inner time step of 1 fs and an outer time step of 4 fs and HMR. For benzene DMTS simulations using the generic model, we needed to reduce the outer time step to 3 fs to avoid instabilities. This confirms that the system specific model is more robust than the generic one.

Figure 3 shows the predicted HFEs compared to the STS reference values. The system-specific model achieved a mean



**Figure 3.** Hydration free energy of small molecules (water, ethanol, benzene, trimethylamine, diethyl sulfide, and acetic acid) using DMTS with system-specific model (blue points) and with small generic model (orange points) compared to the STS result. MAE is 0.091 and 0.103 kcal/mol, RMSE is 0.124 and 0.138 kcal/mol, and  $R^2$  is 0.996 and 0.995 for the system-specific model and the generic model, respectively.

absolute error (MAE) of 0.091 kcal/mol, a root-mean-square error (RMSE) of 0.124 kcal/mol, and a coefficient of determination ( $R^2$ ) of 0.996. The generic foundation model also showed good performance, with an MAE of 0.103 kcal/mol, an RMSE of 0.138 kcal/mol, and an  $R^2$  of 0.995.

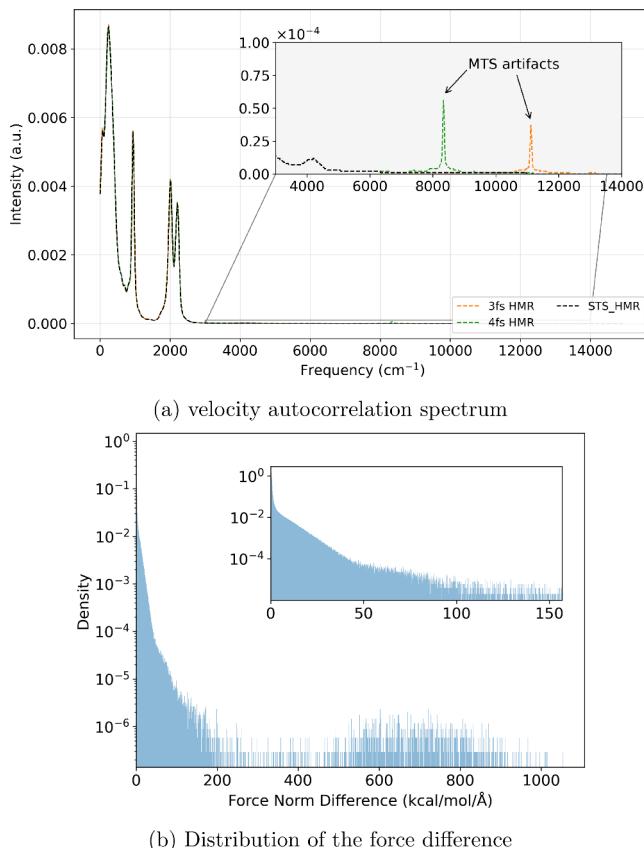
These results demonstrate that the DMTS scheme preserves high accuracy in free energy calculations while benefiting from a significant reduction in the computational cost. They also show that both system-specific and generic models are capable of closely reproducing STS-level results.

## PROTEIN–LIGAND COMPLEXES

To validate our scheme on a biologically relevant system, we performed a 20 ns NVT molecular dynamics simulation of the lysozyme–phenol complex (PDB ID: 4I7L) in explicit water. Because of the chemical diversity of this system, we did not train a system-specific model from scratch. We thus first assessed the ability of the generic model to handle the system and then showed that its performance can be improved via active-learning-driven fine-tuning.

Using an inner time step of 1.75 fs and an outer time step of 3.5 fs with HMR, we obtained stable simulations over 20 ns

with the generic model while preserving the protein's structure and the ligand's binding mode as shown in Figure 5. Note that simulations with a 4 fs outer time step and HMR (which were stable for homogeneous systems) displayed instabilities after 2 ns. The close similarity of the velocity autocorrelation spectra (Figure 4a) between STS and DMTS indicates that these



**Figure 4.** (a) Average velocity autocorrelation spectra of hydrogen atoms for the phenol-lysozyme complex with HMR, comparing STS with MTS 3 and 4 fs. Inset shows zoom in high frequency regions containing MTS integration artifacts. (b) Distribution (in log-scale) of the norm of the force differences for the generic model with respect to the FENNIX-Bio1(M) reference model. The distributions are estimated over the first 500 frames of a phenol-lysozyme in water simulation.

instabilities are not due to the resonance phenomena described above. They instead originate from the presence of "holes" in the potential energy surface of the small model, yielding large differences between its forces and the target ones. This is illustrated in Figure 4b showing the distribution of the norm of the difference of forces between FeNNix-Bio1(M) and the distilled model along a STS trajectory of the protein-ligand system. We observed a dense distribution of values close to 0  $\text{kcal/mol}/\text{\AA}$  and a long-tailed distribution starting from 150  $\text{kcal/mol}/\text{\AA}$ . The latter is associated with large nonphysical force discrepancies yielding the infrequent instabilities that we observe in the DMTS simulation.

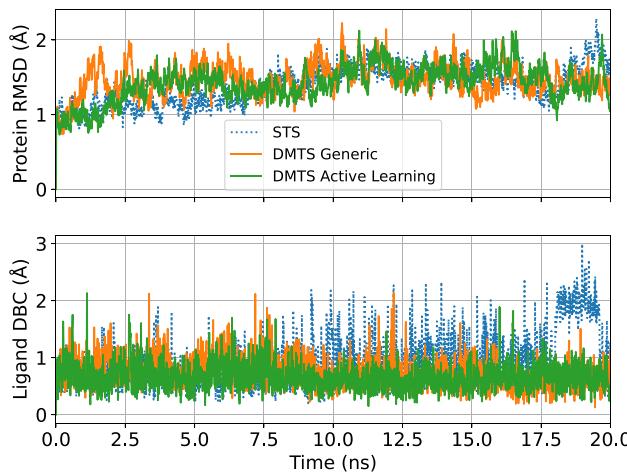
To address this limitation, we introduced two complementary strategies:

**Small message-passing model.** First, we explored enriching the small model architecture with explicit message passing. The intuition is that a more expressive model, with more parameters and deeper local information flow, should exhibit

fewer "holes" in the potential energy surface, thereby reducing the occurrence of pathological configurations. This strategy proved effective: with a single message passing layer, we obtained stable 20 ns simulations using 1 fs/4 fs inner/outer time steps, confirming the improved robustness of the small model (see Figure S2 in Supporting Information). However, the added architectural complexity significantly increased the inference cost, making the approach less attractive for large-scale production simulations (going from 6.54 ns/day for the generic model to 5.95 ns/day for the message-passing model).

**Improving the small model via active learning.** We therefore investigated an alternative strategy that preserved the simplicity of the original model: active-learning-driven fine-tuning. In this approach, we designed an active-learning procedure that automatically detects frames and atoms in which the force difference exceeds an unrealistic threshold of 150  $\text{kcal/mol}/\text{\AA}$ . For such frames, the integrator temporarily reverts to the STS integrator, and local clusters centered on the problematic atoms are added to an adaptive fine-tuning data set. In total, curating the final data set, required approximately 400 ps of simulation (see details in Supporting Information). The generic small model was then refined on this curated data set, systematically improving its robustness in regions of configuration space previously associated with instabilities.

Remarkably, this refinement procedure achieved, without increasing the model's complexity, stable 20 ns simulations using up to 2 fs/4 fs inner/outer time steps. The protein's structure and the ligand's binding mode were preserved as shown in Figure 5. Furthermore, we show in the Supporting Information (Figure S1) that the potential energy is stable along the whole 20 ns simulations in all cases. With this configuration, we reached a production speed of 7.45 ns/day, corresponding to a 2.92 ns speed increase compared to the STS simulations, strongly accelerating production simulations with biological systems. This demonstrates that targeted fine-



**Figure 5.** Time evolution of the protein backbone RMSD and the ligand's Distance to Bound Configuration (DBC)<sup>52</sup> during a 20 ns simulation of the lysozyme-phenol complex in water. Results obtained with the implemented MTS integrator (solid lines) are compared to those from a reference STS simulation (dotted lines). DMTS simulations use either the generic or active learning models combined with the FeNNix-Bio1(M) potential with an internal time step of 1.75 fs and an external time step of 3.5 fs with HMR for the generic model and an internal/external time step of 2/4 fs with HMR for the active learning model.

tuning can rival the stability gains of more expressive architectures while retaining a significantly lower inference cost. Looking ahead, this adaptive strategy could be further enhanced through more sophisticated active-learning schemes.<sup>50,51</sup>

We have introduced a practical and efficient Distilled Multiple Time Step (DMTS) scheme tailored to machine-learned force fields, enabling significant acceleration of molecular dynamics simulations without compromising accuracy for both static and dynamic properties. Our approach, based on a RESPA-like scheme, couples a fast, distilled neural network model with the accurate FeNNix-Bio1(M) reference potential. We demonstrated two complementary distillation strategies: (i) on-the-fly system-specific model distillation and (ii) the use of a lightweight transferable model for broader applicability.

We achieved substantial speed gains (up to 4 in mostly homogeneous systems (bulk water and solvated small molecules) and 2.92 in large protein–ligand complexes) while preserving key physical observables such as radial distribution functions, hydration free energies, diffusion coefficients, and protein–ligand structural properties. Note that this is a first estimation of the computational gains, as the code has yet to be optimized further to handle the dual-level approach more efficiently. All in all, our approach allows us to reach above 7 ns of simulation per day on a single A100 GPU for a realistic protein–ligand complex while preserving the *ab initio*-like accuracy of the FeNNix-Bio1(M) model. In combination with well-established accelerated sampling schemes,<sup>53–55</sup> this enabled truly large-scale simulations with foundation neural network potentials.

Future work will focus on expanding the applicability of this approach through two main avenues. First, randomized time stepping could be explored to further increase the effective time step, while allowing for controlled velocity rescaling as proposed in stochastic RESPA variants, i.e., JUMP integrators.<sup>48</sup> Second, combining the fragment-based active-learning strategy with the generic message passing architecture offers a promising route forward. An active-learning-driven message-passing model could selectively refine unstable regions on the fly, potentially enabling an inner time step beyond 4 fs while offsetting the higher inference cost of the message passing.

## ■ ASSOCIATED CONTENT

### Data Availability Statement

The generic pretrained model can be found on Github at <https://github.com/FeNNol-tools/FeNNol-PMC>

### SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpcllett.5c03720>.

Model architecture and training procedure details. Time evolution of the protein potential energy during 20 ns simulations using STS and DMTS integration methods, generic and active learning models. Time evolution of the protein backbone RMSD and the ligand's DBC during a 20 ns simulation, using STS and DMTS integration methods, generic, active learning and message passing models. (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Authors

Thomas Plé – Sorbonne Université, Laboratoire de Chimie Théorique, UMR 7616 CNRS, 75005 Paris, France;  
Email: [thomas.ple@sorbonne-universite.fr](mailto:thomas.ple@sorbonne-universite.fr)

Jean-Philip Piquemal – Sorbonne Université, Laboratoire de Chimie Théorique, UMR 7616 CNRS, 75005 Paris, France;  
ORCID: [orcid.org/0000-0001-6615-9426](https://orcid.org/0000-0001-6615-9426); Email: [jean-philip.piquemal@sorbonne-universite.fr](mailto:jean-philip.piquemal@sorbonne-universite.fr)

### Authors

Côme Cattin – Sorbonne Université, Laboratoire de Chimie Théorique, UMR 7616 CNRS, 75005 Paris, France

Olivier Adjoua – Sorbonne Université, Laboratoire de Chimie Théorique, UMR 7616 CNRS, 75005 Paris, France

Nicolaï Gouraud – Qubit Pharmaceuticals, Advanced Research Department, 75014 Paris, France; ORCID: [orcid.org/0009-0004-0722-037X](https://orcid.org/0009-0004-0722-037X)

Louis Lagardère – Sorbonne Université, Laboratoire de Chimie Théorique, UMR 7616 CNRS, 75005 Paris, France; ORCID: [orcid.org/0000-0002-7251-0910](https://orcid.org/0000-0002-7251-0910)

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.jpcllett.5c03720>

### Notes

The authors declare the following competing financial interest(s): L. L. and J.-P. P. are shareholders and co-founders of Qubit Pharmaceuticals.

## ■ ACKNOWLEDGMENTS

This work has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 810367), project EMC2 (J.-P. P.). Computations have been performed at IDRIS (Jean Zay) on GENCI Grants: no A0150712052 (J.-P. P.).

## ■ REFERENCES

- (1) Behler, J.; Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **2007**, *98*, 146401.
- (2) Bartók, A. P.; Payne, M. C.; Kondor, R.; Csányi, G. Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons. *Phys. Rev. Lett.* **2010**, *104*, 136403.
- (3) Chmiela, S.; Sauceda, H. E.; Müller, K.-R.; Tkatchenko, A. Towards exact molecular dynamics simulations with machine-learned force fields. *Nat. Commun.* **2018**, *9*, 3887.
- (4) Shakouri, K.; Behler, J.; Meyer, J.; Kroes, G.-J. Accurate neural network description of surface phonons in reactive gas–surface dynamics: N<sub>2</sub> + Ru (0001). *J. Phys. Chem. Lett.* **2017**, *8*, 2131–2136.
- (5) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chemical Science* **2017**, *8*, 3192–3203.
- (6) Schütt, K. T.; Kindermans, P.-J.; Sauceda, H. E.; Chmiela, S.; Tkatchenko, A.; Müller, K.-R. SchNet: a continuous-filter convolutional neural network for modeling quantum interactions. *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Red Hook, NY, USA **2017**, 992–1002.
- (7) Unke, O. T.; Chmiela, S.; Gastegger, M.; Schütt, K. T.; Sauceda, H. E.; Müller, K.-R. SpookyNet: Learning force fields with electronic degrees of freedom and nonlocal effects. *Nat. Commun.* **2021**, *12*, 7273.

(8) Gasteiger, J.; Becker, F.; Günemann, S. Gemnet: Universal directional graph neural networks for molecules. *Advances in Neural Information Processing Systems* **2021**, *34*, 6790–6802.

(9) Musaelian, A.; Batzner, S.; Johansson, A.; Sun, L.; Owen, C. J.; Kornbluth, M.; Kozinsky, B. Learning local equivariant representations for large-scale atomistic dynamics. *Nat. Commun.* **2023**, *14*, 579.

(10) Grisafi, A.; Ceriotti, M. Incorporating long-range physics in atomic-scale machine learning. *J. Chem. Phys.* **2019**, *151*, 204105.

(11) Devereux, C.; Smith, J. S.; Huddleston, K. K.; Barros, K.; Zubatyuk, R.; Isayev, O.; Roitberg, A. E. Extending the Applicability of the ANI Deep Learning Molecular Potential to Sulfur and Halogens. *J. Chem. Theory Comput.* **2020**, *16*, 4192–4202.

(12) Zubatyuk, R.; Smith, J. S.; Leszczynski, J.; Isayev, O. Accurate and transferable multitask prediction of chemical properties with an atoms-in-molecules neural network. *Science Advances* **2019**, *5*, eaav6490.

(13) Grisafi, A.; Nigam, J.; Ceriotti, M. Multi-scale approach for the prediction of atomic scale properties. *Chemical Science* **2021**, *12*, 2078–2090.

(14) Kabylda, A.; Frank, J. T.; Suárez-Dou, S.; Khabibrakhmanov, A.; Medrano Sandonas, L.; Unke, O. T.; Chmiela, S.; Müller, K.-R.; Tkatchenko, A. Molecular Simulations with a Pretrained Neural Network and Universal Pairwise Force Fields. *J. Am. Chem. Soc.* **2025**, *147*, 33723–33734.

(15) Unke, O. T.; Stöhr, M.; Ganscha, S.; Unterthiner, T.; Maennel, H.; Kashubin, S.; Ahlin, D.; Gastegger, M.; Sandonas, L. M.; Berryman, J. T.; Tkatchenko, A.; Müller, K.-R. Biomolecular dynamics with machine-learned quantum-mechanical force fields trained on diverse chemical fragments. *Science Advances* **2024**, *10*, eadn4397.

(16) Batatia, I.; Kovacs, D. P.; Simm, G.; Ortner, C.; Csányi, G. MACE: Higher order equivariant message passing neural networks for fast and accurate force fields. *Advances in Neural Information Processing Systems* **2022**, *35*, 11423–11436.

(17) Kovács, D. P.; Moore, J. H.; Browning, N. J.; Batatia, I.; Horton, J. T.; Pu, Y.; Kapil, V.; Witt, W. C.; Magdäu, I.-B.; Cole, D. J.; Csányi, G. MACE-OFF: Short-Range Transferable Machine Learning Force Fields for Organic Molecules. *J. Am. Chem. Soc.* **2025**, *147*, 17598–17611.

(18) Plé, T.; Lagardère, L.; Piquemal, J.-P. Force-field-enhanced neural network interactions: from local equivariant embedding to atom-in-molecule properties and long-range effects. *Chemical Science* **2023**, *14*, 12554–12569.

(19) Plé, T.; Adjoua, O.; Benali, A.; Posenitskiy, E.; Villot, C.; Lagardère, L.; Piquemal, J.-P. A Foundation Model for Accurate Atomistic Simulations in Drug Design. *ChemRxiv* **2025**, DOI: [10.26434/chemrxiv-2025-f1hgn-v4](https://doi.org/10.26434/chemrxiv-2025-f1hgn-v4).

(20) Benali, A.; et al. Pushing the Accuracy Limit of Foundation Neural Network Models with Quantum Monte Carlo Forces and Path Integrals. *arXiv* **2025**, DOI: [10.48550/arXiv.2504.07948](https://doi.org/10.48550/arXiv.2504.07948).

(21) MacKerell, A. D. Empirical Force Fields for Proteins: Current Status and Future Directions; *Annual Reports in Computational Chemistry*; Elsevier, 2005; Vol. 1, Chapter 7, pp 91–102, DOI: [10.1016/S1574-1400\(05\)01007-8](https://doi.org/10.1016/S1574-1400(05)01007-8).

(22) Tuckerman, M.; Berne, B. J.; Martyna, G. J. Reversible multiple time scale molecular dynamics. *J. Chem. Phys.* **1992**, *97*, 1990–2001.

(23) Zhou, R.; Harder, E.; Xu, H.; Berne, B. Efficient multiple time step method for use with Ewald and particle mesh Ewald for large biomolecular systems. *J. Chem. Phys.* **2001**, *115*, 2348–2358.

(24) Lagardère, L.; Aviat, F.; Piquemal, J.-P. Pushing the limits of multiple-time-step strategies for polarizable point dipole molecular dynamics. *J. Phys. Chem. Lett.* **2019**, *10*, 2593–2599.

(25) Liberatore, E.; Meli, R.; Rothlisberger, U. A Versatile Multiple Time Step Scheme for Efficient ab Initio Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2018**, *14*, 2834–2842.

(26) Jaffrelot Inizan, T.; Plé, T.; Adjoua, O.; Ren, P.; Gokcan, H.; Isayev, O.; Lagardère, L.; Piquemal, J.-P. Scalable hybrid deep neural networks/polarizable potentials biomolecular simulations including long-range effects. *Chemical Science* **2023**, *14*, 5438–5452.

(27) Fu, X.; Musaelian, A.; Johansson, A.; Jaakkola, T.; Kozinsky, B. Learning interatomic potentials at multiple scales. *arXiv preprint* **2023**, No. arXiv:2310.13756, DOI: [10.48550/arXiv.2310.13756](https://doi.org/10.48550/arXiv.2310.13756).

(28) Osadchey, R.; Nam, K.; Cui, Q. Toward Improving Multiple Time Step QM/MM Simulations with  $\Delta$ -Machine Learning. *J. Phys. Chem. B* **2025**, *129*, 10451–10466.

(29) Mouvet, F.; Browning, N. J.; Baudin, P.; Liberatore, E.; Rothlisberger, U. Machine learning-enhanced multiple time-step ab initio molecular dynamics. *J. Chem. Phys.* **2025**, *163*, 184115.

(30) Biesiadecki, J. J.; Skeel, R. D. Dangers of Multiple Time Step Methods. *J. Comput. Phys.* **1993**, *109*, 318–328.

(31) Ma, Q.; Izaguirre, J.; Skeel, R. Verlet-I/R-RESPA/Impulse is Limited by Nonlinear Instabilities. *SIAM Journal on Scientific Computing* **2003**, *24*, 1951–1973.

(32) Hinton, G.; Vinyals, O.; Dean, J. Distilling the knowledge in a neural network. *arXiv preprint* **2015**, No. arXiv:1503.02531, DOI: [10.48550/arXiv.1503.02531](https://doi.org/10.48550/arXiv.1503.02531).

(33) Plé, T.; Adjoua, O.; Lagardère, L.; Piquemal, J.-P. FeNNol: An efficient and flexible library for building force-field-enhanced neural network potentials. *J. Chem. Phys.* **2024**, *161*, 042502.

(34) Lagardère, L.; Jolly, L.-H.; Lipparini, F.; Aviat, F.; Stamm, B.; Jing, Z. F.; Harger, M.; Torabifard, H.; Cisneros, G. A.; Schnieders, M. J.; Gresh, N.; Maday, Y.; Ren, P. Y.; Ponder, J. W.; Piquemal, J.-P. Tinker-HP: a massively parallel molecular dynamics package for multiscale simulations of large complex systems with advanced point dipole polarizable force fields. *Chemical Science* **2018**, *9*, 956–972.

(35) Adjoua, O.; Lagardère, L.; Jolly, L.-H.; Durocher, A.; Very, T.; Dupays, I.; Wang, Z.; Inizan, T. J.; Célerse, F.; Ren, P.; Ponder, J. W.; Piquemal, J.-P. Tinker-hp: Accelerating molecular dynamics simulations of large complex systems with advanced point dipole polarizable force fields using gpus and multi-gpu systems. *J. Chem. Theory Comput.* **2021**, *17*, 2034–2053.

(36) Gou, J.; Yu, B.; Maybank, S. J.; Tao, D. Knowledge distillation: A survey. *International journal of computer vision* **2021**, *129*, 1789–1819.

(37) Eastman, P.; Behara, P. K.; Dotson, D. L.; Galvelis, R.; Herr, J. E.; Horton, J. T.; Mao, Y.; Chodera, J. D.; Pritchard, B. P.; Wang, Y.; et al. SPICE, A Dataset of Drug-like Molecules and Peptides for Training Machine Learning Potentials. *Scientific Data* **2023**, *10*, 11.

(38) Eastman, P.; Pritchard, B. P.; Chodera, J. D.; Markland, T. E. Nutmeg and SPICE: Models and Data for Biomolecular Machine Learning. *J. Chem. Theory Comput.* **2024**, *20*, 8583–8593.

(39) Leimkuhler, B.; Matthews, C. Rational construction of stochastic numerical methods for molecular sampling. *Appl. Math. Res. Express* **2012**, DOI: [10.1093/amrx/abs010](https://doi.org/10.1093/amrx/abs010).

(40) Leimkuhler, B.; Matthews, C. Robust and efficient configurational molecular sampling via Langevin dynamics. *J. Chem. Phys.* **2013**, *138*, 174102.

(41) Rackers, J. A.; Wang, Z.; Lu, C.; Laury, M. L.; Lagardère, L.; Schnieders, M. J.; Piquemal, J.-P.; Ren, P.; Ponder, J. W. Tinker 8: software tools for molecular design. *J. Chem. Theory Comput.* **2018**, *14*, 5273–5289.

(42) Feenstra, K. A.; Hess, B.; Berendsen, H. J. C. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *J. Comput. Chem.* **1999**, *20*, 786–798.

(43) Morrone, J. A.; Markland, T. E.; Ceriotti, M.; Berne, B. J. Efficient multiple time scale molecular dynamics: Using colored noise thermostats to stabilize resonances. *J. Chem. Phys.* **2011**, *134*, 014103.

(44) Leimkuhler, B.; Margul, D. T.; Tuckerman, M. E. Stochastic, resonance-free multiple time-step algorithm for molecular dynamics with very large time steps. *Mol. Phys.* **2013**, *111*, 3579–3594.

(45) Margul, D. T.; Tuckerman, M. E. A Stochastic, Resonance-Free Multiple Time-Step Algorithm for Polarizable Models That Permits Very Large Time Steps. *J. Chem. Theory Comput.* **2016**, *12*, 2170–2180.

(46) Albaugh, A.; Tuckerman, M. E.; Head-Gordon, T. Combining Iteration-Free Polarization with Large Time Step Stochastic-Isokinetic Integration. *J. Chem. Theory Comput.* **2019**, *15*, 2195–2205.

(47) Gouraud, N.; Bris, P. L.; Majka, A.; Monmarché, P. HMC and Underdamped Langevin United in the Unadjusted Convex Smooth Case. *SIAM/ASA Journal on Uncertainty Quantification* **2025**, *13*, 278–303.

(48) Gouraud, N.; Lagardère, L.; Adjoua, O.; Plé, T.; Monmarché, P.; Piquemal, J.-P. Velocity Jumps for Molecular Dynamics. *J. Chem. Theory Comput.* **2025**, *21*, 2854–2866.

(49) Lagardère, L.; Maurin, L.; Adjoua, O.; El Hage, K.; Monmarché, P.; Piquemal, J.-P.; Hénin, J. Lambda-ABF: Simplified, Portable, Accurate, and Cost-Effective Alchemical Free-Energy Computation. *J. Chem. Theory Comput.* **2024**, *20*, 4481–4498.

(50) van der Oord, C.; Sachs, M.; Kovács, D. P.; Ortner, C.; Csányi, G. Hyperactive Learning for Data-Driven Interatomic Potentials. *npj Computational Materials* **2023**, *9*, 168.

(51) Kulichenko, M.; Barros, K.; Lubbers, N.; Li, Y. W.; Messerly, R.; Tretiak, S.; Smith, J. S.; Nebgen, B. Uncertainty-Driven Dynamics for Active Learning of Interatomic Potentials. *Nature Computational Science* **2023**, *3*, 230–239.

(52) Salari, R.; Joseph, T.; Lohia, R.; Hénin, J.; Brannigan, G. A streamlined, general approach for computing ligand binding free energies and its application to GPCR-bound cholesterol. *J. Chem. Theory Comput.* **2018**, *14*, 6560–6573.

(53) Célerse, F.; Inizan, T. J.; Lagardère, L.; Adjoua, O.; Monmarché, P.; Miao, Y.; Derat, E.; Piquemal, J.-P. An Efficient Gaussian-Accelerated Molecular Dynamics (GaMD) Multilevel Enhanced Sampling Strategy: Application to Polarizable Force Fields Simulations of Large Biological Systems. *J. Chem. Theory Comput.* **2022**, *18*, 968–977.

(54) Jaffrelot Inizan, T.; Célerse, F.; Adjoua, O.; El Ahdab, D.; Jolly, L.-H.; Liu, C.; Ren, P.; Montes, M.; Lagarde, N.; Lagardère, L.; Monmarché, P.; Piquemal, J.-P. High-resolution mining of the SARS-CoV-2 main protease conformational space: supercomputer-driven unsupervised adaptive sampling. *Chemical Science* **2021**, *12*, 4889–4907.

(55) Ansari, N.; Jing, Z. F.; Gagelin, A.; Hédin, F.; Aviat, F.; Hénin, J.; Piquemal, J.-P.; Lagardère, L. Lambda-ABF-OPES: Faster Convergence with High Accuracy in Alchemical Free Energy Calculations. *J. Phys. Chem. Lett.* **2025**, *16*, 4626–4634.