Will it blend? Mixing numerical and machine-learned physics quantities for accurate on-the-fly surrogate modeling

Michael Tynes[1,2][0000−0002−5007−1056], Kyle Chard[1,2][0000−0002−7370−4805], Ian Foster[1,2][0000−0003−2129−5269], and Logan Ward[2][0000−0002−1323−5939]

[1] Department of Computer Science, University of Chicago, Chicago IL 60637, USA
mtynes@uchicago.edu chard@uchicago.edu
[2] Data Science and Learning Division, Argonne National Laboratory, Lemont IL, USA lward@anl.gov

Abstract. Learning and deploying inexpensive replacements for computationally expensive subroutines "on-the-fly" (OTF) during a dynamic simulation offers potential advantages and unique drawbacks. In OTF learning, a machine learned surrogate function is trained to replace a target subroutine in a simulation as the dynamics evolve. The advantages of OTF learning include reducing simulation error and model training costs, but the weaknesses include the possibility of introducing artacts when adaptively updating the physics of the simulation over time. Here, we enhance an existing control system, Proxima, which ensures that surrogates are used appropriately in time-independent state-sampling based simulations, to time-dependent dynamical simulations by introducing a "blending" procedure that hides discontinuities when transitioning between original subroutine and surrogate. Our new control system, Proxima+Blend, produces a blend of the surrogate and target functions according to the relationship between error and an uncertainty signal observed as the dynamic simulation evolves. We show that while the original control system can shorten application runtime and accurately capture some macro-scale observables of a molecular dynamics simulation, the addition of blending is necessary to avoid unphysical dynamics at shorter time and length scales and to correct observables derived from these dynamics. Proxima+Blend delivers a 1.5x speedup over use of the target subroutine while producing solutions within 5% error in dynamical quantities, while the original Proxima algorithm has up to 80% error. Our implementation of Proxima+Blend can be deployed by simply replacing the existing subroutine with a wrapper that includes a machine learning approach for surrogate training along with specifying control and uncertainty signals for the simulation of interest.

Keywords: Machine learning · dynamic simulations · uncertainty quantification

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97632-2_19

1 Introduction

Machine learning (ML) surrogates can speed up physical simulations by orders of magnitude, granting access to previously unavailable length and time scales, but optimal methods for constructing such surrogates and integrating them into simulations are still being developed. A classic route is to invest resources into gathering data and training ahead of time, keeping the surrogate fixed during execution [25]. Keeping the model fixed simplifies deployment at the risk of poor performance if the simulation samples states unseen in the training data. On-the-fly (OTF) learning strategies mitigate this risk by continually updating surrogates during the course of a simulation [9, 27]. However, while promising, OTF learning comes with its own uncertainties and challenges.

Existing OTF methods enhance a simulation by updating a surrogate over the course of a simulation by gradually training it on data acquired during execution. While simple in concept, running a simulation with a constantly changing surrogate presents subtle challenges. Reliably training machine learning models for a particular physics code is, of course, a challenge. Balancing the computational cost differences between target and surrogate requires non-trivial system design, but such challenges have been overcome in previous cases [7, 20]. Even with a model with sufficient accuracy and a deployment strategy in hand, knowing whether the outcome of the simulation is correct is not assured.

For one, metrics to predict the quality of ML predictions are often unreliable [1, 21]. The inability to detect when a simulation has begun to explore states different than those encountered in training inhibits a scientist's ability to know when they should be skeptical of an ML-driven simulation. At best, a strategy based on quality metric involves introducing a control parameter (e.g., a threshold on an uncertainty metric) which must be carefully selected and continually re-evaluated. Below, we further explore a strategy to automate auditing and adjusting such control parameters introduced by Zamora *et al.* [30].

Another, yet-unaddressed question with OTF learning is whether replacing a target function with a mutable, learning object can lead to issues in the enveloping application. In the ideal case, the machine learned surrogate is indistinguishable from the original function so any algorithms which employ it should be unaffected. However, achieving a perfect surrogate is impossible. While pleasingly straightforward, we posit that naively switching a surrogate on or off will lead to unphysical results in dynamical simulations by causing rapid changes to variables that are expected by numerical integrators to vary smoothly. Established methods which blend multiple fidelities of models together, such as the classic QM/MM technique, employ methods to blend differing levels across spatial dimensions [28]. Below, we demonstrate blending in time is needed as well.

In this work, we present a control system strategy which simplifies integrating surrogate models into dynamic simulations without modifying the original algorithms. We build on the existing Proxima approach [30], which uses a linear controller to determine when a surrogate model should be used. We propose that smoothly adjusting convex combinations of ML surrogate predictions and

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97632-2_19

physics calculations will reduce unphysical results which stem from turning a surrogate on or off immediately. We demonstrate that our new method, Proxima+Blend, fixes instability issues found in molecular dynamics simulations and achieves accurate simulation outputs without sacrificing the accelerations provided by surrogates.

2 Methods

Our approach developed here, Proxima+Blend, extends an existing method for replacing functions with machine-learned surrogates, Proxima [30]. Proxima is an OTF method that uses a control system to ensure that a surrogate function is used only when it meets user-prescribed accuracy bounds by adaptively tuning an uncertainty quantification (UQ) threshold.

2.1 Existing Proxima implementation

Here we formally describe the implementation of the existing Proxima algorithm 30, an abstract description of the simulation, surrogate and target functions, uncertainty quantification, and a control strategy. The goal of the algorithm is to maximize use the surrogate while ensuring that the average of the errors between the surrogate and target functions adhere to user-specified bound.

We let x_t be the state of the system under simulation at time t; F be a computationally expensive target function that is used to compute $y_t = F(x_t)$, a physical quantity in this system that determines the simulation evolution; $\hat{y}_t = S(x_t)$ be an ML surrogate estimate of this quantity; $\epsilon_t = |y_t - \hat{y}_t|$ be the error between S and F at t; ϵ_{bound} be the user-specified error bound; and UQ be a measure of the uncertainty of S; UQ_t^{thr} be the current uncertainty threshold. The UQ metric is commonly chosen to the variance of y_t among an ensemble of surrogate models or the distance between x_t and all points included in the training set for S.

Multiple computations occur at each step of the simulation:

1. *Reliability check* The UQ metric is computed and compared against the current threshold. If $UQ < UQ_t^{thr}$, then declare the surrogate unreliable. Optionally, we also declare a certain fraction of surrogate evaluations unreliable regardless of the outcome of the UQ test.
2. *Retraining* If the training set has grown by more than some threshold amount since the last training and the surrogate is declared reliable, the surrogate (S) is retrained.
3. *Evaluation* If the surrogate is determined to be reliable, S is used to compute y_t; otherwise, F is used and the resultant (x_t, y_t) point is stored.
4. *Threshold Update* If $UQ \geq UQ_t^{thr}$, UQ^{thr} is updated according to the error between the surrogate and target $\epsilon_t = |y_t - \hat{y}_t| = |S(x_t) - F(x_t)|$

Proxima generally, adjusts UQ_t^{thr} to be smaller (more conservative) when the error (ϵ_t) is high and larger (less conservative) when the error is low. The

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97632-2_19

original Proxima implementation used a linear controller to adjust UQ_t^{thr},

$$UQ_{t+1}^{thr} = UQ_t^{thr} - \alpha_t(\epsilon_t - \epsilon_{target}) \tag{1}$$

where α_t is a regression coefficient fit to the accumulated dataset of UQ and observed errors.

2.2 Proxima+Blend implementation

To extend Proxima to use in dynamical simulations, we introduce a new approach that smoothly transitions between the target and surrogate functions over multiple timesteps. The blend is controlled by a mixing parameter, $\lambda \in [0, 1]$, which defines the amount of surrogate S and target function F used in evaluating the output:

$$\hat{y}_t^{\text{blend}} = \lambda_t S(x_t) + (1 - \lambda_t)F(x_t). \tag{2}$$

When $\lambda_t = 1$, the simulation does not require evaluating the target function.

Whereas the original implementation of Proxima transitions between two values of λ ($\{0, 1\}$), our new approach Proxima+Blend varies transitions smoothly over several timesteps. The number of timesteps it takes for a full transition from $\lambda = 0$ to $\lambda = 1$ is a user-supplied parameter n. As detailed in Algorithm 1, the degree of mixing is represented by a blending index (i) that varies between 0 and n by ± 1 depending on whether Proxima's control algorithm decides the surrogate could be used. (The original Proxima algorithm is where $n = 1$.) The degree of mixing λ is finally determined transforming i/n through a smoothing function, f, which has gradients tending to zero at arguments 0 and 1, ensuring that blending away from full use of ML (S) or the target function (F) occurs slowly. In this work, we use

$$f(z) = -cos(\pi x) + 1, \tag{3}$$

but any sigmoidal function taking on values from 0 to 1 could be used.

2.3 Computational evaluation overview

We implement Proxima and Proxima+Blend for atomistic simulations, where the system under study is a set of atoms distributed in 3-dimensional space. Specifically, we consider molecular dynamics (MD), where the positions and momenta of each atom are evolved over time by integrating a $6N$-dimensional set of ordinary differential equations which are primarily Newton's Laws of Motion. The expensive step in propagating the MD simulation is often computing the forces acting on each atom as a function of the relative positions of all other atoms. The gold standard for computing the forces on atoms is *ab initio* quantum mechanical calculations, such as those based on Density Functional Theory (DFT). The run duration of molecular dynamics simulations is typically dicated by the force compuations, with more accurate forces taking more time to compute and having worse algorithmic complexity, making the force computation a popular target for ML surrogate modeling [2, 3, 18, 24].

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97632-2_19

Algorithm 1 Algorithm for blending physics and machine-learned quantities

if $UQ < UQ_t^{thr}$ **then**
  $i \leftarrow i + 1$
else
  $i \leftarrow i - 1$
end if
$i \leftarrow \text{clip}(i, 0, n)$
$\lambda_t \leftarrow f(i/n)$
if i = n **then**
  $\hat{y} \leftarrow S(x_t)$
else
  $\hat{y} \leftarrow \lambda_t S(x_t) + (1 - \lambda_t)F(x_t)$
end if
return $\hat{y}$

We implemented Proxima and Proxima+Blend as a `Calculator` class in the Atomic Simulation Environment (ASE) [16]. ASE provides a common Python interface to various tools for atomic simulation, including many tools to perform *ab initio* calculations. ASE also provides classes for the numerical integration in MD, which depend on Calculator classes to produce forces at each simulation step. Our ASE `Calculator` class has as attributes two other ASE `Calculator` objects–one for S and one for F—and implements the logic from Proxima+Blend to take convex combinations of their results, based on the varying the threshold UQ^{thr}, and to update λ as presented in Algorithm 1.

Our implementation is available online on github [3].

Further details on simulations For our MD simulations, we used the ASE interface to CP2K [13] to perform DFT force evaluations. These calculations used the local density approximation functional and Gaussian plane wave basis set (LDA/GPW) as our target physics F. (Full details of our DFT configuration are available on GitHub[4]).

For our surrogate, we used the ANI neural network architecture [24], a widely used neural network approach for predicting energies and forces for atomistic systems, as implemented in the TorchANI package [11]. ANI takes as input a set of atomic species and coordinates, and predicts the potential energy of the atomic system and the forces acting on each atom, with the forces obtained by automatic differentiation of the energy contribution of each atom with respect to the atomic positions. We optimize a weighted sum of Huber loss terms for the energy, forces, and stressses. Stress is also determined by automatic differentiation of the predicted energy, in this case with respect to strain applied to each coordinate in each spatial direction. In our experiments, we trained our networks

[3] https://github.com/globus-labs/cascade
[4] https://github.com/globus-labs/cascade/blob/main/cascade/files/cp2k-lda-template.inp

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97632-2_19

for 32 epochs using the Adam optimizer with initial learning rate 10^{-3}. We used an ensemble of two ANI networks each trained using a different bootstrapped sample of available training data, with the intent of using differences between ensemble members as an uncertainty metric. More details on hyperparameters and our fitting procedure can be found on GitHub[5].

We configured PROXIMA and PROXIMA+BLEND to control the maximum error between the target DFT forces (our instance of F) and those predicted by the surrogate (S). The input control signal is the maximum deviation in the forces predicted by any model from the mean over all models. The error (ϵ_t) and the regression coefficient (α_t) are determined using averages over the last eight timesteps for which the target model was used.

Our simulations examine crystalline silicon at various temperatures. Specifically, our system under study is a lattice of 64 Silicon atoms under periodic boundary conditions, with a single atom removed to create a vacancy, allowing us to study the diffusion of this vacancy throughout the system: an important determinant of the mechanical properties of solids. We performed simulations at constant volume and temperature (NVT) and constant pressure and temperature (NPT) for various temperatures and pressures.

3 Results

The goals of our experiments are to (1) explore whether switching between a numerical routine and a machine-learned surrogate distorts the dynamics of a molecular dynamics simulation, and (2) evaluate strategies for mitigating distortions.

3.1 On-the-fly learning in a MD simulation

We start by evaluating whether OTF learning techniques result in correct behavior in a classic dynamics problem: constant temperature and pressure molecular dynamics (often called NPT, for fixed **n**umber of particles, **p**ressure, and **t**emperature). We choose a standard approach using a variant of the popular Nosé–Hoover thermostat, which controls temperature through a coupling of the system with a virtual heat bath, and of the Parniello-Raman barostat, which controls pressure through the equilibrium between the stress state of the atoms and a virtual "piston" that represents external pressure [19]. The piston term includes momentum, which leads to oscillations around equilibrium and implies a strong dependence of the evolution of a system on its history. This history dependence creates a challenge that could expose issues in replacing a function with a surrogate.

The goal of our computation is to estimate the density of silicon at 800K. We do so by starting with a 63-atom cell of Si atoms at 0K, adding random velocities

[5] https://github.com/globus-labs/cascade/blob/main/2_proxima/0_run-serial-proxima.py

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97632-2_19

corresponding to an initial temperature of 800K, and evolving Newton's laws of motion and the NPT thermostat and barostat over 2048 timesteps (a total of 2048 fs). The density of the system is related to the average volume over the last 512 timesteps, after giving the thermostat and barostat time to equilibrate the pressure and temperature. This computation is simple enough to solve without machine learning acceleration, and using the target numerical routine at every step yields an equilibrium density of 2.166 g/cc.

Introducing OTF learning involves replacing the calculation of the forces used in Newton's laws of motion and the stress used in the NPT with a surrogate. We start with a simple strategy: first gather 1024 data points from F before training an ensemble of neural networks S, allow the Proxima and Proxima+Blend algorithms to decide when to use F, S, or a blend of each to evolve the dynamics without further training. We set the error target ϵ_{bound} to 0.9 eV/Å(well above the convergence threshold for the target method), add stochasticity such that the target method at least 20% of the time.

Viewing the resulting calculations at a coarse level, OTF learning has only minor differences in the simulation outcomes, giving an equilibrium density of 2.163 g/cc, only a 0.11% error with respect to the reference calculations, while yielding a 1.57x speedup (3.41 hrs / MD run vs. 2.18 hours). We note that the runtimes are dominated by force evaluations and that we observed average timestep durations of 4 seconds when using DFT forces and 0.2 seconds when using ML forces (so roughly a 20x speedup), though with considerable variation depending on atomic configuration and the recent runtime history of the DFT routine.

Viewing the calculations at a finer level, the errors between surrogate and target ϵ_t have a median below the user-supplied bound of ϵ_{bound}, that is the error is below the bound slightly more than half the time (Figure 1). The mean of ϵ_t is, however, above the bound. The discrepancy between the mean and median is driven by high-error outliers. On one hand, such outliers are to be expected because the variation in forces on atoms in MD simulations are very large, while on the other hand, this does present an opportunity to develop a more advanced control strategy. We will not pursue this further in this work, since the majority of errors are within the threshold and other problems must be solved first to use Proxima in dynamical simulations.

While Proxima accurately captures the density of the system, there are problems in the small-scale dynamics. We expect the volume of the system to fluctuate over time with oscillation periods much larger than the simulation timestep (1 fs). Instead, there are oscillations of with periods on the order of 10 fs (Figure 2). The outcome of the simulation is correct on one metric, but alarmingly wrong on another.

3.2 Introducing blending into on-the-fly ML-driven dynamics

Switching abruptly from the target to surrogate function creates a rapid, non-smooth change in the system over time, which is not only unphysical but could lead to long-lasting effects in the dynamics of the system. Thus, we propose

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97632-2_19

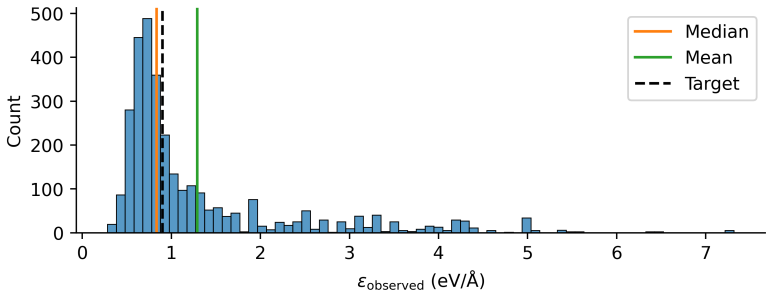

Fig. 1. A histogram of observed errors between the surrogate and target functions, their mean and median, and the target error from an example trajectory

a gradual transition between one function and another. Detailed in §2.2, this method slowly introduces the surrogate over time when uncertainty is below the threshold established by the Proxima algorithm, and likewise slowly reduces surrogate usage when the uncertainty is above the threshold.

The goal of our blending method is to mitigate the changes in the dynamics introduced by rapid switching, and this is what we see in our results. Figure 2 shows the volume oscillation over a segment of one of these trajectories. We can see that once Proxima begins switching between the surrogate and the target a high-frequency oscillation is introduced to the Proxima trajectory which is not present in the target-only trajectory and the amplitude of the lower frequency oscillation increases, both indicating a departure from the expected dynamics of the system caused by rapid switching. When using Proxima+Blend, no such oscillations are present, indicating that our blending method has the intended effect.

The insight gained from inspecting an individual trajectory above is further supported by quantifying the high-frequency oscillations. Rapid oscillations will be associated with large time derivatives, so we calculate the maximum absolute gradient of the volume over time across all trajectories; in doing so we see in Figure 3 that Proxima has much larger gradients than the trajectories generated by the target method, while Proxima+Blend is nearly identical to the target method, indicating a major reduction in rapid oscillations.

While we expect blending to reduce artifacts introduced by switching between target and surrogate, we expected it to be slower than the un-blended Proxima due to the possibility of more calls to the target method being performed. However, we observed that the speedup of Proxima+Blend over DFT is only slightly less than that of Proxima without blending, as shown in Figure 4, indicating that the additional evaluations of F when the UQ is under the threshold but $\lambda < 1$ are infrequent enough to make Proxima+Blend a useful tool for accelerating calculations.

The promise of the blending method is that it does not require altering the original algorithms. There are, indeed, algorithms for molecular dynamics which

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97632-2_19

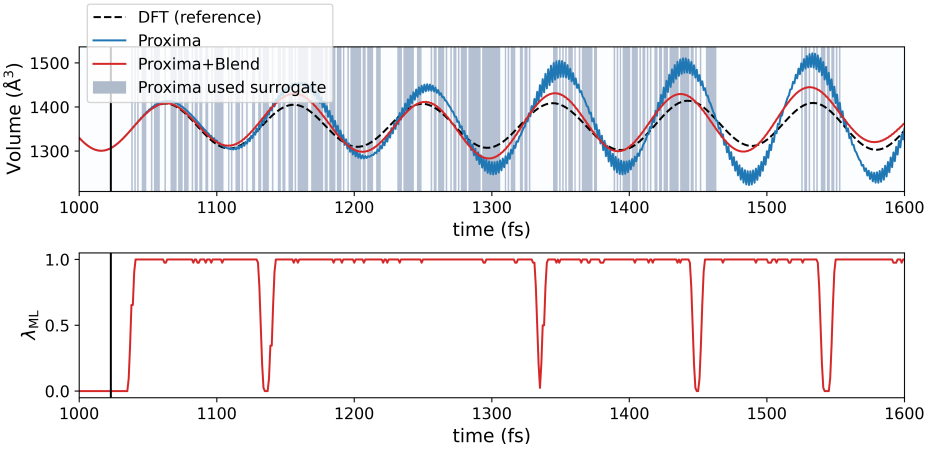

Fig. 2. The top panel shows the fluctuation of the volume of the simulation cell over time across conditions from a selected velocity configuration. The dashed line is derived from the trajectory that used DFT forces at every time step. The blue and red lines are derived from the Proxima and Proxima+Blend trajectories, respectively. The vertical black line indicates when the first surrogate was trained and could be used by Proxima or Proxima+Blend. The blue shaded regions indicate when Proxima (without blending) used ML to advance the trajectory. The bottom pane shows the value of λ with Proxima+Blend.

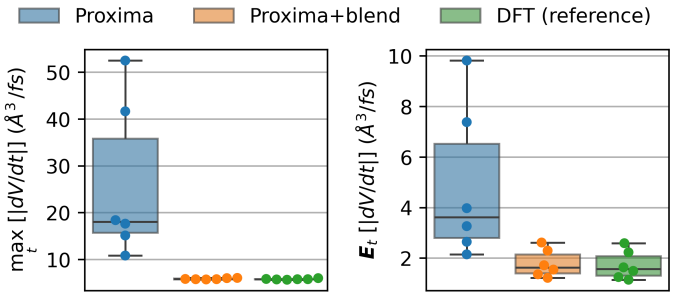

Fig. 3. A box plot of the maximum gradient in volume with respect to time (Å^3/fs), derived from the trajectories generated by DFT, Proxima, and Proxima+Blend.

are less sensitive to the abrupt changes of introducing surrogates. We found that molecular dynamics performed with Berendsen barostat does not exhibit the fluctuations, perhaps because the volume changes depend only on the pressure at the present timestep [4]. Our method eliminates the need to carefully examine whether OTF learning is appropriate for a new simulation process. Validation and adjustment of the blending timescales are necessary, but there are fewer limits to which algorithms can tolerate changing between original and surrogate on-the-fly.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97632-2_19

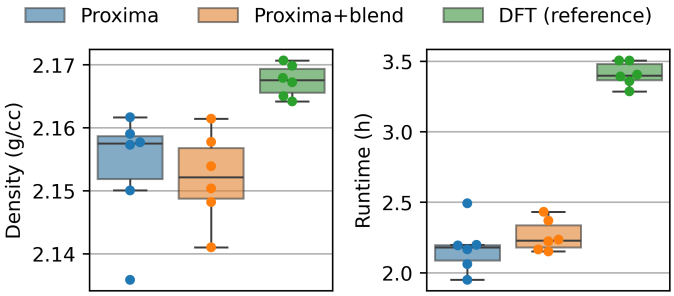

Fig. 4. A box plot of the runtimes of the simulations by condition.

3.3 Blending corrects fine-trained dynamics

Eliminating the unphysical oscillations of cell volume using Proxima+Blend is a positive result; however, volume is only one of the 373 degrees of freedom in our molecular dynamics calculation. We further explore the reliability of OTF learning by studying the dynamics of the remaining degrees of freedom (atom positions, momenta) by measuring how often the system transitions between different states. Specifically, we perform a simulation where one atom is removed from an otherwise regular crystal lattice and then we measure the frequency that another atom "hops" into the newly-vacant site (known as a vacancy). Vacancy hopping is controlled by interactions and motion of many individual atoms and, thus, requires accurate dynamics to model correctly.

To examine the effects of switching and blending on vacancy diffusion, we performed a set of NVT (that is, fixed **n**umber, **v**olume, and **t**emperature) molecular dynamics simulations at the equilibrium volume determined from previous NPT simulations. NVT is often used when examining diffusion due to effects of the barostat on such bevavior. We used a higher temperature (1573K) to increase the hopping rate. The key metric of our calculation is the time between hops, which we detect by determining when the neighbors associated with each atom change. We detect neighbors by counting the number of atoms within a distance slightly larger than the equilibrium bond distance in Si.

Our results show that hop-rates derived from Proxima+Blend agree with those generated by the target function F within 5%, while those from Proxima (without blending) disagree by more than 80%, further indicating that our addition of blending removes artifacts and makes simulations more accurate. As shown in Table 1, using Proxima to control the surrogate estimates a hop rate nearly twice as fast as the surrogate-free, DFT-only calculation.

A remaining challenge highlighted by the hop-rate study is that the erroneous hop rate is not simple to find. As with density calculation, Proxima's controller achieved the target level of error between surrogate and target function. The unphysical dynamics, however, were not apparent until comparing hop rate with a full-fidelity calculation. The ability to perform a full-fidelity calculation is un-

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97632-2_19

available in many cases and techniques for directly relating error in a specific subroutine to larger scale observables are limited. As such, we recommend users of OTF learning — with Proxima or otherwise — to carefully examine the effect of algorithm parameters on their target outcome. Even with such a limitation, adding Proxima+Blend reduces the chance for unphysical behavior and improves agreement with full-fidelity dynamics compared to Proxima.

Table 1. Estimates of the Si vacancy hop rate estimated from trajectories generated by DFT, Proxima, and Proxima+Blend, respectively

Method	Hop rate (SE) [1/ps=THz]
DFT-only	2.68 (0.0834)
Proxima	5.06 (0.114)
Proxima+Blend	2.80 (0.0852)

4 Related Work

4.1 Machine-learned Surrogates for Expensive Computations

Developing cheap approximates for expensive calculations is a widespread activity in modern computational sciences and is becoming even more effective with advancements in machine learning. Within our focus domain, electronic-scale physics, cheap approximates for computing atomic forces have nearly a century of practice [17] and are undergoing a renaissance with advanced machine learning techniques [3]. The machine learning models create surrogates with many more degrees of freedom than those created by humans, which are then able to recover more nuances of the function being approximated.

The scope of surrogates for simulations extends well past learning subroutines which compute forces on atoms. Tools which approximate specific subroutines are growing in impact for climate simulations, in particular [22, 23]. Methods for learning a faster set of dynamic equations rather than just replacements for evaluating specific are also emerging [6, 15].

The tradeoff for the intricate surrogates available from machine learning is the lack of a clear understanding of the limitations of the model. As such, approaches for estimating the uncertainty of the model are of key importance. Techniques range from evaluating differences between models trained with varied training sets or directly training a confidence signal that will be output with every prediction [26]. Regardless of method, the quality of the confidence interval must be assumed to be imperfect and must be calibrated to produce meaningful results [14].

4.2 Offline Training of Surrogates

The conventional practice for machine-learned surrogates is to separate training a surrogate from using it to solve a science problem. The workflow for creating

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97632-2_19

a surrogate includes *enumerating* the space of potential inputs to a subroutine, *evaluating* many to produce a training set, and *training* the surrogate.

Enumeration, evaluation, and training are often coupled. Modern practices for making surrogates use an active learning procedure which alternates between training a model and using the model to identify which potential inputs to evaluate [12]. The active learning procedure may even involve enumerating new points by solving example problems using the surrogate model [5].

4.3 On-the-fly Training of Surrogates

Early approaches to OTF learning on atomistic simulations can be found in Refs 8–10. These approaches decompose a molecular dynamics simulation into regions where quantum-level accuracy is and is-not required, and train spline-based parametric surrogates on-the-fly in the high-fidelity regions. Instead of blending surrogate and target methods to address discontinuities, these works make use of made use of a rollback and interpolation procedures where timesteps around model updates are repeated and rerun with the new model, or in the case of Refs 8, 9 rerun with a version of their surrogate interpolated between their new and old one. The target method was never used to run dynamics directly.

Similar spatially decomposed + OTF surrogate construction approaches have been employed more recently, including in the massively parallel approach described by Ref 7.

Recently an application called FLARE has been developed for OTF training of Gaussian-process based atomistic surrogates [27, 29]. These model forms have the advantages of needing relatively small training datasets and separating data and parameter uncertainty, allowing the parameter uncertainty to be used to trigger surrogate updates when necessary.

5 Conclusion

We have demonstrated that control methods can ensure accurate outcomes while introducing machine-learned surrogates into dynamics simulations, provided that the transition to using the faster surrogate is smooth. We modify an existing algorithm Proxima [30], that enforces that surrogates are used only when appropriate, by adding a technique to adjust the blend between original and surrogate subroutines rather than toggling abruptly. This new approach, Proxima+Blend, eliminates unphysical oscillations of volume in constant pressure molecular dynamics simulations and corrects the timescales of state transitions in constant-volume dynamics. Importantly, Proxima+Blend required no changes in the dynamical equations. We envision that providing such autonomous control systems will enable the reduction in computational cost by providing a reliable cost/accuracy tradeoff in more software across computational sciences.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97632-2_19

Acknowledgements

LW was supported by the Office of Defense Nuclear Nonproliferation Research and Development within the U.S. Department of Energy's National Nuclear Security Administration. MT was supported by U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Computational Science Graduate Fellowship under Award Number(s) DE-SC0023112. This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

References

1. Anderson, G.J., Gaffney, J.A., Spears, B.K., Bremer, P.T., Anirudh, R., Thiagarajan, J.J.: Meaningful uncertainties from deep neural network surrogates of large-scale numerical simulations (2020)
2. Batatia, I., Benner, P., ..., C., Csányi, G.: A foundation model for atomistic materials chemistry (Mar 2024). https://doi.org/10.48550/arXiv.2401.00096, http://arxiv.org/abs/2401.00096, arXiv:2401.00096 [physics]
3. Behler, J.: Perspective: Machine learning potentials for atomistic simulations. The Journal of Chemical Physics **145**(17) (Nov 2016). https://doi.org/10.1063/1.4966192, http://dx.doi.org/10.1063/1.4966192
4. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A., Haak, J.R.: Molecular dynamics with coupling to an external bath. The Journal of Chemical Physics **81**(8), 3684–3690 (Oct 1984). https://doi.org/10.1063/1.448118, http://dx.doi.org/10.1063/1.448118
5. Bernstein, N., Csányi, G., Deringer, V.L.: De novo exploration and self-guided learning of potential-energy surfaces. npj Computational Materials **5**(1), 1–9 (Oct 2019). https://doi.org/10.1038/s41524-019-0236-6, https://www.nature.com/articles/s41524-019-0236-6, publisher: Nature Publishing Group
6. Brunton, S.L., Proctor, J.L., Kutz, J.N.: Discovering governing equations from data by sparse identification of nonlinear dynamical systems. Proceedings of the National Academy of Sciences **113**(15), 3932–3937 (Mar 2016). https://doi.org/10.1073/pnas.1517384113, http://dx.doi.org/10.1073/pnas.1517384113
7. Caccin, M., Li, Z., Kermode, J.R., De Vita, A.: A framework for machine-learning-augmented multiscale atomistic simulations on parallel supercomputers. International Journal of Quantum Chemistry **115**(16), 1129–1139 (2015). https://doi.org/10.1002/qua.24952, https://onlinelibrary.wiley.com/doi/abs/10.1002/qua.24952, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qua.24952

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97632-2_19

8. Csányi, G., Albaret, T., Moras, G., Payne, M.C., Vita, A.D.: Multiscale hybrid simulation methods for material systems. Journal of Physics: Condensed Matter **17**(27), R691 (Jun 2005). https://doi.org/10.1088/0953-8984/17/27/R02, https://dx.doi.org/10.1088/0953-8984/17/27/R02
9. Csányi, G., Albaret, T., Payne, M.C., De Vita, A.: "Learn on the Fly": A Hybrid Classical and Quantum-Mechanical Molecular Dynamics Simulation. Physical Review Letters **93**(17), 175503 (Oct 2004). https://doi.org/10.1103/PhysRevLett.93.175503, https://link.aps.org/doi/10.1103/PhysRevLett.93.175503
10. De Vita, A., Car, R.: A Novel Scheme for Accurate Md Simulations of Large Systems. MRS Online Proceedings Library **491**(1), 473–480 (Dec 1997). https://doi.org/10.1557/PROC-491-473, https://doi.org/10.1557/PROC-491-473
11. Gao, X., Ramezanghorbani, F., Isayev, O., Smith, J.S., Roitberg, A.E.: TorchANI: A Free and Open Source PyTorch-Based Deep Learning Implementation of the ANI Neural Network Potentials. Journal of Chemical Information and Modeling **60**(7), 3408–3415 (Jul 2020). https://doi.org/10.1021/acs.jcim.0c00451, https://doi.org/10.1021/acs.jcim.0c00451, publisher: American Chemical Society
12. Guo, Y., Nath, P., Mahadevan, S., Witherell, P.: Active learning for adaptive surrogate model improvement in high-dimensional problems. Structural and Multidisciplinary Optimization **67**(7) (Jul 2024). https://doi.org/10.1007/s00158-024-03816-9, http://dx.doi.org/10.1007/s00158-024-03816-9
13. Kühne, T.D., Iannuzzi, M.., Hutter, J.: Cp2k: An electronic structure and molecular dynamics software package-quickstep: Efficient and accurate electronic structure calculations. The Journal of Chemical Physics **152**(19) (2020)
14. Kuleshov, V., Fenner, N., Ermon, S.: Accurate uncertainties for deep learning using calibrated regression (2018). https://doi.org/10.48550/ARXIV.1807.00263, https://arxiv.org/abs/1807.00263
15. Kurth, T., Subramanian, S., Harrington, P., Pathak, J., Mardani, M., Hall, D., Miele, A., Kashinath, K., Anandkumar, A.: Fourcastnet: Accelerating global high-resolution weather forecasting using adaptive fourier neural operators. In: Proceedings of the Platform for Advanced Scientific Computing Conference. p. 1–11. PASC '23, ACM (Jun 2023). https://doi.org/10.1145/3592979.3593412, http://dx.doi.org/10.1145/3592979.3593412
16. Larsen, A.H., Mortensen, J.J.., Jacobsen, K.W.: The atomic simulation environment—a Python library for working with atoms. Journal of Physics: Condensed Matter **29**(27), 273002 (Jun 2017). https://doi.org/10.1088/1361-648X/aa680e, https://dx.doi.org/10.1088/1361-648X/aa680e, publisher: IOP Publishing
17. Lennard-Jones, J.E.: Cohesion. Proceedings of the Physical Society **43**(5), 461–482 (Sep 1931). https://doi.org/10.1088/0959-5309/43/5/301, http://dx.doi.org/10.1088/0959-5309/43/5/301
18. Lubbers, N., Smith, J.S., Barros, K.: Hierarchical modeling of molecular energies using a deep neural network. The Journal of Chemical Physics **148**(24), 241715 (Jun 2018). https://doi.org/10.1063/1.5011181, http://arxiv.org/abs/1710.00017, arXiv:1710.00017 [stat]
19. Melchionna, S., Ciccotti, G., Lee Holian, B.: Hoover npt dynamics for systems varying in shape and size. Molecular Physics **78**(3), 533–544 (Feb 1993). https://doi.org/10.1080/00268979300100371, http://dx.doi.org/10.1080/00268979300100371
20. Partee, S., Ellis, M., Rigazzi, A., Shao, A.E., Bachman, S., Marques, G., Robbins, B.: Using machine learning at scale in numerical simulations with smartsim: An application to ocean climate modeling. Journal of Computational Science **62**, 101707 (Jul 2022). https://doi.org/10.1016/j.jocs.2022.101707, http://dx.doi.org/10.1016/j.jocs.2022.101707

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97632-2_19

21. Peterson, A.A., Christensen, R., Khorshidi, A.: Addressing uncertainty in atomistic machine learning. Physical Chemistry Chemical Physics **19**(18), 10978–10985 (2017). https://doi.org/10.1039/c7cp00375g, http://dx.doi.org/10.1039/C7CP00375G
22. Rasp, S., Pritchard, M.S., Gentine, P.: Deep learning to represent subgrid processes in climate models. Proceedings of the National Academy of Sciences **115**(39), 9684–9689 (Sep 2018). https://doi.org/10.1073/pnas.1810286115, http://dx.doi.org/10.1073/pnas.1810286115
23. Sastry, V.K., Maulik, R., Rao, V., Lusch, B., Renganathan, S.A., Kotamarthi, R.: Data-Driven Deep Learning Emulators for Geophysical Forecasting, p. 433–446. Springer International Publishing (2021). https://doi.org/10.1007/978-3-030-77977-1_35, http://dx.doi.org/10.1007/978-3-030-77977-1_35
24. 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 **8**(4), 3192–3203 (Mar 2017). https://doi.org/10.1039/C6SC05720A, https://pubs.rsc.org/en/content/articlelanding/2017/sc/c6sc05720a, publisher: The Royal Society of Chemistry
25. Smith, J.S., Nebgen, B., Lubbers, N., Isayev, O., Roitberg, A.E.: Less is more: Sampling chemical space with active learning. The Journal of Chemical Physics **148**(24), 241733 (May 2018). https://doi.org/10.1063/1.5023802, https://doi.org/10.1063/1.5023802
26. Tran, K., Neiswanger, W., Yoon, J., Zhang, Q., Xing, E., Ulissi, Z.W.: Methods for comparing uncertainty quantifications for material property predictions. Machine Learning: Science and Technology **1**(2), 025006 (2020)
27. Vandermause, J., Torrisi, S.B., Batzner, S., Xie, Y., Sun, L., Kolpak, A.M., Kozinsky, B.: On-the-fly active learning of interpretable Bayesian force fields for atomistic rare events. npj Computational Materials **6**(1), 1–11 (Mar 2020). https://doi.org/10.1038/s41524-020-0283-z, https://www.nature.com/articles/s41524-020-0283-z, publisher: Nature Publishing Group
28. Warshel, A., Levitt, M.: Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. Journal of Molecular Biology **103**(2), 227–249 (May 1976). https://doi.org/10.1016/0022-2836(76)90311-9, http://dx.doi.org/10.1016/0022-2836(76)90311-9
29. Xie, Y., Vandermause, J., Ramakers, S., Protik, N.H., Johansson, A., Kozinsky, B.: Uncertainty-aware molecular dynamics from Bayesian active learning for phase transformations and thermal transport in SiC. npj Computational Materials **9**(1), 1–8 (Mar 2023). https://doi.org/10.1038/s41524-023-00988-8, https://www.nature.com/articles/s41524-023-00988-8, publisher: Nature Publishing Group
30. Zamora, Y., Ward, L., Sivaraman, G., Foster, I., Hoffmann, H.: Proxima: accelerating the integration of machine learning in atomistic simulations. In: Proceedings of the ACM International Conference on Supercomputing. pp. 242–253. ICS '21, Association for Computing Machinery, New York, NY, USA (Jun 2021). https://doi.org/10.1145/3447818.3460370, https://dl.acm.org/doi/10.1145/3447818.3460370

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97632-2_19