Ab Initio Generalized Langevin Equations

Pinchen Xie  
Program in Applied and Computational Mathematics,  
Princeton University, Princeton, NJ 08544, USA

Roberto Car  
Department of Chemistry, Department of Physics,  
Program in Applied and Computational Mathematics,  
Princeton Institute for the Science and Technology of Materials,  
Princeton University, Princeton, NJ 08544, USA

Weinan E  
AI for Science Institute, Beijing, China,  
Center for Machine Learning Research and School of Mathematical Sciences,  
Peking University, Beijing, China  
(Dated: November 15, 2022)

We propose an approach for learning accurately the dynamics of slow collective variables from atomistic data obtained from ab-initio quantum mechanical theory, using generalized Langevin equations (GLE). The force fields, memory kernel, and noise generator are constructed within the Mori-Zwanzig formalism under the constraint imposed by the fluctuation-dissipation theorem. Combined with Deep Potential Molecular Dynamics (DeePMD) and density functional theory, this GLE approach opens the door to carry out first-principles multi-scale modeling for a variety of systems. Here, we demonstrate this capability with a study of the dynamics of twin domain walls in ferroelectric lead titanate. The importance of memory effects is illustrated by the fact that while ab-initio GLE agrees well with molecular dynamics at near-equilibrium conditions, Markovian Langevin dynamics underestimates the rate of rare events by several orders of magnitude.

Developing accurate and reliable meso-scale physical models has been a long-standing problem in science with implications to many areas including material modeling and computational biology [1–5]. In this long scientific pursuit, two ideas stand out. The first is the Landau approach for the free energy in terms of phenomenological order parameters. The second is the Mori-Zwanzig formalism for the dynamics of coarse-grained variables derived from the microscopic degrees of freedom.

The key concept in Landau models is the order parameter. It is an observable or a set of observables that in a symmetry breaking phase transition take nonzero value(s) in the ordered phase and vanish in the disordered phase. The microscopic origin of the order parameter is usually ignored, while the free energy is expressed by a physically motivated series expansion in terms of the order parameters and their gradients. The resultant effective model is generally referred to as Landau-Ginzburg theory. It has been applied successfully to a variety of phenomena, including superconductivity [6], ferromagnetism [7], ferroelectricity [8], etc. Dynamical equations can also be constructed on phenomenological grounds. Well-known examples include the Landau-Lifshitz equation for the evolution of the magnetization [9], the Allen-Cahn and the Cahn-Hilliard equations for the dynamics of phase separation [10 11], more generally the phase field models [12] and the phase-field-crystal models [13–15] for a variety of problems.

For the purpose of getting physical insight, Landau’s approach is a stroke of genius, but from a more practical perspective there have been concerns on the quantitative accuracy of the Landau models, especially when they are used in regimes out of equilibrium. This state of affairs bears some similarity with the state of the molecular dynamics (MD) field before ab initio MD was proposed.

The Mori-Zwanzig formalism is a general methodology for constructing effective models for any set of collective variables (CVs) and can in principle be made accurate. The idea is to project the original dynamics on the space of the CVs. This introduces two important new effects: memory, because the dynamics of the CVs is generally non-Markovian, and noise, due to the initial condition for the variables eliminated in the projection process. These effects are difficult to model accurately. As a consequence, in practical applications, one typically resorts to approximate forms such as the generalized Langevin equation (GLE) with simple memory and noise terms to describe the effective dynamics. Incidentally, the Landau dynamical equations can be regarded as an even more drastic approximation of the projected dynamics, using physical intuition to suggest the form of the effective model [16].

In recent years, machine learning has emerged as a powerful tool for developing accurate and reliable effective models in related contexts [17–28]. A notable success was achieved in the MD field, where neural network-based inter-atomic potentials made possible to perform very large-scale simulations with ab initio accuracy [29 30]. In this paper we introduce a machine learning methodology to construct reliable GLE models.
from atomistic models derived from DFT. In our approach, the memory is time invariant with finite time scale, while the noise satisfies the constraint imposed by the fluctuation dissipation theorem. We call the schemes derived in this way ab initio generalized Langevin equation (AIGLE) models, as they can be derived from first-principles atomistic models having quantum mechanical accuracy. GLEs can be rigorously derived within the Mori-Zwanzig formalism only if the microscopic Hamiltonian is quadratic [31]. Thus, AIGLE is not the most general ansatz for CV dynamics. It is, however, sufficiently flexible and general to connect microscopic and mesoscopic dynamics in the same spirit in which semi-local DFT bridges quantum mechanics and atomistic modeling. The great potential of AIGLE in material and molecular modeling stems from the fact that it allows us to take full consideration of memory and noise effects. In this paper, we explore its application in the simulation of domain wall dynamics in ferroelectric materials. We present an ab initio multi-scale study of nucleation-driven ferroelectric domain motion, in which the AIGLE formulation is based on models for the interatomic interactions and the electric dipoles derived from quantum mechanics in the context of DFT.

RESULTS

Machine learning-based GLE model

The starting point is a fully atomistic model for the molecular dynamics. We define some collective variables (CVs) in terms of the microscopic degrees of freedom. Our task is to eliminate the remaining degrees of freedom and obtain an accurate dynamic model for the CVs. The main idea of the MZ formalism is to project the dynamics of the CVs into two orthogonal subspaces, the subspace of the CVs and the subspace of the eliminated variables. It is relatively straightforward to express such a projection process at the level of a master equation, as is often done in applications of the MZ formalism. In practice, one has to make drastic simplifications and start with some prescribed ansatz. A popular ansatz, say for an atomistic system, is the following

\[ M \dddot{x}(t) = -\nabla G(x(t)) + F(t) + \int_0^t ds MK(s) \frac{dx(t-s)}{dt} + R(t). \]

Here the column vector \( x \) represents the CVs, \( M \) the effective mass matrix, \( G(x) \) the free energy surface, \( F \) the external driving forces, \( K \) the memory kernel matrix, and \( R \) the noise. It is convenient to define \( \dot{v} = \frac{dx}{dt} \), \( a = \ddot{x} \), \( F = -\nabla G(x) + F \), and \( v^T \) the transpose of \( v \). We require \( \langle R(t)v^T(0) \rangle = 0 \) and the orthogonality condition \( \langle R(t)v^T(0) \rangle = 0 \). Here and in the rest of the paper, the brackets \( \langle \cdots \rangle \) indicate average over the equilibrium ensemble at \( t = 0 \). The generalization of the second fluctuation-dissipation theorem (2FDT) [52] prescribes that \( \langle v(0)v^T(0) \rangle K^T(s) = -\langle (M^{-1}R)(0)(M^{-1}R)^T(s) \rangle \) at equilibrium.

In our numerical approach one learns the non-Markovian GLE of Eq. (1) from MD trajectories of \( x \). The scheme can use but does not require a predetermined free energy surface [26,33,38] — all the terms in the GLE can be learned from adequate trajectory data. The resulting GLE satisfies numerically the 2FDT for the equilibrium ensemble. But the model can be extended to simulate near-equilibrium dynamics.

Numerous studies have been reported on data-driven parameterizations of the GLE [39–47], the Markovian Langevin equations [48–50] and the far-from-equilibrium GLE [51,52]. Most relevant to our work are recent investigations based on time-series regression [42] and extended dynamics [47]. Ref. [42] modeled a generic stochastic differential system by forecasting the time series via a nonlinear autoregressive model. We take a step further in this regard by modeling the noise term in the GLE with deep learning. Ref. [47] modeled GLEs by incorporating 2FDT with auxiliary variables and hierarchical rational approximations. As in many other works, Ref. [47] assumes that the force fields are predetermined. Our unified GLE model is more flexible in this regard while respecting 2FDT. We also allow the co-determination of a nearly steady driving force for simplifying the study of near-equilibrium dynamics such as driven domain motions.

Though our approach can be extended to general non-linear dynamics beyond 2FDT, in this paper we limit the scope to near-equilibrium atomistic systems. Our goal is to demonstrate the applicability of the proposed method to simulated MD data. As a proof of concept, in the following, we demonstrate our approach with a one-dimensional CV.

For learning from MD data it is convenient to transform the GLE ansatz of Eq. (1) into a discrete form. We assume that the memory kernel is homogeneous, i.e. free of dependence on position [53] or time origin [54]. Let \( \Delta t \) be the time step of the discretized GLE, \( t = 0 \) be an arbitrary starting point and \( t = n \Delta t \) be the current time. We use \( f_{(n)} \) as an abbreviation for any time-dependent function \( f(n \Delta t) \). Our ansatz for a one-dimensional GLE then reads

\[ m a_{(n)} = -\nabla G(x_{(n)}) + F_{(n)} + \sum_{s=0}^{n-1} m K_{(s+s+\frac{1}{2})} v_{(n-s-\frac{1}{2})} \Delta t + R_{(n)}. \]

The GLE is propagated by the leapfrog scheme:

\[ v_{(n+\frac{1}{2})} = v_{(n-\frac{1}{2})} + a_{(n)} \Delta t, \]
\[ x_{(n+1)} = x_{(n)} + v_{(n+\frac{1}{2})} \Delta t. \]

The leapfrog scheme is second order, similar to the velocity Verlet algorithm commonly implemented in MD codes [55,56]. This setup allows direct alignment with MD data when \( \Delta t \) is equal to an integer multiple of the time step \( \delta t \) of the MD simulation. We require \( \Delta t \) to be small enough for stability.
The free energy $G$ in Eq. (2) can be parameterized empirically, using for example a polynomial ansatz, or more generally, it can be represented by a neural network ansatz when dealing with high-dimensional CVs. The noise term $\{ R_n \}$ in Eq. (2) is modeled by the generalized autoregressive (GAR) model

$$ R_n = \sum_{k=1}^{m_A} \phi(k) R_{n-k} + \mu_n + \sigma_n w(n) $$

(4)

where $\{ \phi(k) \}$ are the Yule-Walker linear autoregressive parameters [57] and

$$ \mu_n = \mu(R_{n-1}, \cdots, R_{n-m_A}), $$

$$ \sigma_n = \sigma(R_{n-1}, \cdots, R_{n-m_A}) $$

(5)

are non-linear functions depending on the history of noise. The non-Markovianity of the noise is due to insufficient separation of time scales. The residual noise $w(n)$ is expected to be standard Gaussian white noise. It represents the uncorrelated part of the noise on the scale of $\Delta t$. In our approach $\mu_n$ and $\sigma_n$ are the output of a deep neural network taking $R_{n-1}, \cdots, R_{n-m_A}$ as arguments. The GAR model reduces to the standard AR($m_A$) model [58] when $\mu_n$ and $\sigma_n$ are constants. In practice, we find that GAR obviously outperforms AR in reducing $w(n)$ to an almost ideal white noise upon training with MD data.

Eqs. (2-5) together compose the AIGLE model. The parameters in Eq. (2) and Eq. (4) are learned from MD data under the following assumptions. MD is supposed to be at, or close to, equilibrium due to interactions with stochastic environments that act like heat baths. Memory effects on such open systems are assumed to extend to a finite time interval specified by the integer $m_K$, i.e., $K(s) = 0$ for $s > m_K$. The learning procedure is then divided into three steps, whose details are given in the Method section. The first step consists of training $G$ and $K$ using equilibrated MD data without invoking Eq. (3). The orthogonality condition $\langle R_n v(0) \rangle = 0$ is reinforced iteratively and the noise is constrained by the maximum likelihood principle. In the second step, we optimize the GAR model using a maximum likelihood loss function, with a protocol in which the extracted $\{ R_n \}$ act as the data. Then, we verify that the residual noise is almost gaussian white noise, and test the stationarity of the GAR model numerically. Upon completing the second step, the GLE is fully determined for equilibrium systems. We can then incorporate extra near-equilibrium MD data, if needed, under a perturbing driving force $F$. At this stage we train only the parameters representing $G$ and $F$ with the maximum likelihood principle.

**Validation with solvable model**

In this section we study a toy model with a known memory kernel [39] in the MZ framework under thermal equilibrium conditions. The model consists of a one-dimensional harmonic chain of $N + 1$ identical particles (see Method section for details). The chain has one free end and one fixed end. The CV $x$ is chosen to be the displacement of the free end away from its equilibrium position.

We acquire MD data by simulating an $N = 100$ chain connected to a Langevin thermostat, modelling a gas/solid interface at $k_B T = 1$. We coarse-grain the data by downsampling. Then we train AIGLE with a free energy $G = \frac{1}{2} \omega_0^2 x^2$ where $\omega_0$ is unknown. The details are described in the Method section. The outcome of the training is reported in Fig. 1. The GAR model successfully reduces $w$ to almost white noise with delta-like normalized autocorrelation function (NACF), as shown in Fig. 1(a). An obvious serial correlation would be present if we used an AR model instead. Both $R$ and $w$ are normally distributed on the dataset. Fig. 1(b) shows that the optimized $K(\tau)$ obeys 2FDT and agrees well with the theoretical result $K^{MZ}(\tau) = -J_1(2\tau)/\tau$ corresponding to an isolated infinite chain. The discrepancy near $\tau = 0$ is due to the inclusion of white noise from the Langevin thermostat. Fig. 1(c) compares the velocity autocorrelation from the MD dataset and the simulated AIGLE. They display minor disagreement for $\tau > 10$. This is the error caused by substituting $w(n)$ with the ideal white noise generator when propagating AIGLE. The same effect can be seen in Fig. 1(a) on the slight discrepancy.
Application to modelling ab initio domain motion

In this section we study a prototypical problem of ferroelectric domain switching — the field-driven motion of a twin 180° domain wall in epitaxial PbTiO₃. Fig. 2 shows a sketch of the domain wall and the crystal structure of PbTiO₃. We let the local order parameter of this system be the effective electric dipole moment (local dipole) $p_j$ associated to each Ti-centered elementary cell $j$ of PbTiO₃, represented by the yellow arrows in Fig. 2. The microscopic definition of $p_j$ in the context of Kohn-Sham DFT is given in the Method section.

For epitaxially strained tetragonal PbTiO₃ polarized along the [001] direction near room temperature ($T = 300K$), the ferroelectric domains have narrow 180° domain walls. The CV for describing continuously the size difference between the +z domain and the −z domain can be defined by $\alpha = \sum_j \text{tanh} \left( \frac{p_j}{p_*} \right)$, where $p_*$ is tuned to ensure that $\alpha$ reproduces well the direct counting of the polarized local dipoles.

The domain motion of PbTiO₃ is known to be glassy when the external electric field $E = E_z$ is weak. Microscopically, the domain motion is initiated by the nucleation and growth of flat nuclei at the domain wall. The velocity of domain motion is phenomenologically described by $\frac{d\alpha}{dt} = v \propto e^{-\frac{\alpha}{E_c}}$. For $E > E_c$, the glassy dynamics due to a succession of rare events is gradually replaced by continuous motion. The domain velocity becomes flow-like, i.e., $v \propto (E-E_c)^\beta$, and can be modeled directly with MD simulations. For $E < E_c$ and $E \approx E_c$, the rare event leading to domain motion is called “domain wall creep”. In this regime, ab initio MD modeling is difficult due to the long time scale (1ns~1ms) of the creeping events. To cope with the time scale of the domain wall creeps, previous studies combined coarse-grained Monte Carlo with the phenomenological Merz’s law and Landau-Ginzburg-Devonshire free energy models. AIGLE allows us to avoid empirical parameterizations and/or phenomenological models. We focus on the $E < E_c$ and $E \approx E_c$ regimes, in which the local dipoles of PbTiO₃ respond linearly to electric fields, and metastable domains can not be reversed in bulk material.

We carry out electric field-driven MD simulations of PbTiO₃ by adopting the Deep Potential (DP) model for the interatomic interactions, and an effective Born charge (BC) model to describe the local dipoles. The DP model is trained on SCAN-DFT data with typical potential energy errors of less than 5meV/atom. The BC model is trained on Berry-phase polarization data consistently calculated with SCAN-DFT. The details are given in the Method section. In the MD simulations, the atomistic degrees of freedom equilibrate quickly with the environment, and the loss of detailed balance is mostly associated to the CV describing domain motion. We generate NP, T-MD trajectories for $\alpha$ in the supercell shown in Fig. 2 in presence of a homogeneous electric field $0 < E < 3$mV/A along the $\hat{z}$ direction. The $x$, $y$ dimensions of the supercell are fixed to match the experimental lattice constant $a = 3.91\AA$. The $z$ dimension is barostatted at constant pressure. The setup is close to that of PbTiO₃ thin films deposited on SrRuO₃/SrTiO₃ substrates. We do not address complications that arise when comparing to experiment, such as the role played by defects, but we comment briefly on finite size effects and the details of MD simulations in the Method section.

Atomistic simulations suggest that the dynamics of $\alpha$ on the nanosecond time scale resembles that of 1D Brownian motion in a tilted periodic potential. On the picosecond time scale, Fourier analysis of $\alpha$ reveals damped harmonic peaks associated to atomistic optical modes, including the fast ones that are irrelevant to domain dynamics. Hence, $\alpha$ should not be used directly for training AIGLE. For this purpose, we use the CV $x$ defined by coarse-graining $\alpha$ along the time axis through truncated Gaussian filtering (see Method section for de-
FIG. 3. (a) The NACF of the noise compared to the rescaled memory kernel \( \tilde{K}(\tau) = K(\tau) \frac{C_{\text{data}}^{AIGLE}(0) + C_{\text{data}}^{\text{MD}}(\Delta t)}{2K(\Delta t/2)} \). (b) The velocity NACF. The inset compares an MD trajectory to a simulated trajectory under \( E = 2\text{mV/A} \). (c) The tilted free energy profile corresponding to two periods of \( U(x) \). \( U(x) \) (blue line) is compared to the metadynamics results (blue plus). The inset shows they agree very well around the basin. (d) The domain wall velocities \( v_D \). Each GLE data (blue dot) is from a 0.5\( \mu \)s long AIGLE simulation. Each MD data (orange dot) is from averaging over one hundred 0.1ns long MD trajectories. The vertical dashed line is at \( E = E^* \). The inset shows the natural logarithm of \( v_D \) as a function of \( 1/E \) for the MD data (orange), the AIGLE (blue), and the Markovian Langevin equation (green) derived from AIGLE. The dashed red line is the best fit of Merz’s law \( \frac{C_{\text{data}}^{AIGLE}}{D} \approx \frac{C_{\text{data}}^{\text{MD}}}{D} \) with the AIGLE data. The best fit of Merz’s law \( E_0 = 27.0\text{mV/A} \) with MD data is not plotted. (e) Simulated displacement of the CV under different \( E \). The dashed red line is at \( 2\text{FDT} \) at metastable equilibrium. The inset shows the natural logarithm of \( v_D \) for the MD data (orange), the AIGLE and from MD are plotted as a function of \( 1/E \). The domain wall velocities \( v_D \) well with each other, with minor discrepancies originating from the small difference between \( C_{\text{data}}^{AIGLE} \) and \( C_{\text{data}}^{\text{MD}} \), respectively from the MD data and from the GLE agree very well with each other, with minor discrepancies originating from the small difference between \( C_{\text{data}}^{AIGLE} \) and \( C_{\text{MD}}^{\text{MD}} \). (f) The scaled hysteresis loop for \( t_p = 1/f = 100\text{\mu}s \) and different \( l \). The “EXP1” experimental data \( [70] \) is measured with the same \( t_p \) and unknown \( l \). The “EXP2” experimental data \( [73] \) is measured with \( l \approx 144\mu\text{m} \) and unknown \( t_p \). The experimental remnant polarization is 94\( \mu \)C/cm\(^2\) \( [70] \) and 96\( \mu \)C/cm\(^2\) \( [73] \) respectively. Our atomistic model gives 72\( \mu \)C/cm\(^2\).

tails). The Gaussian filter is used to suppress irrelevant optical modes. The remaining vibrational modes contribute to memory and noise in the AIGLE formalism (Eq. [2]). Suppression of fast atomic motions allows a large timestep \( \Delta t = 10\text{fs} \). The free energy surface for \( x \) is parameterized by the periodic function \( G(x) = mU(x) = mU_b \tanh(k(1 - \cos \omega(x - x_0))) \). \( m \) is a scalar mass, \( U_b \), the barrier height, is predetermined with metadynamics \([34]\), since it is hard to fit it accurately in the near-equilibrium region without enhanced sampling. The external force in Eq. [2] is represented by \( F(x) = mpE \) assuming linear response regime. \( k, \omega, x_0 \) and \( p \) are free parameters.

We first train AIGLE using MD trajectories with \( E = 2\text{mV/A} \). Under this condition, the MD is at metastable equilibrium on the time scale of the simulation. Then we retrain \( G(x) \) and \( F(x) \) with MD trajectories under different \( E \in [2.0, 2.4]\text{mV/A} \). In this dataset, \( x \) can violate detailed balance through domain motions. The details of the training are given in the Method section.

The resultant noise NACF and the rescaled memory kernel \( \tilde{K} \) are plotted in Fig. 3(a), showing satisfactory agreement with 2FDT at metastable equilibrium. Fig. 3(b) shows that the velocity NACFs extracted respectively from the MD data and from the GLE agree well with each other, with minor discrepancies originating from the small difference between \( C_{\Delta t=5}^{\text{data}} \) and \( C_{\Delta t=5}^{\text{sim}} \) for \( \tau \geq 0.5\text{ps} \) (see Fig. 3(c)).

The optimized free energy surface is shown in Fig. 3(c). The metastability ceases to exist when the tilted energy profile becomes monotonic at \( E^* \approx 2.8\text{mV/A} \). Hence the near-equilibrium region appropriate for the application of AIGLE should be roughly \( E \leq E^* \), in order for the lifetime of metastable equilibrium to be much longer than the relaxation time of atomic vibrational modes. To better visualize this phenomenology, we simulate the AIGLE for a dense grid of electric field values in the interval \([0, 3.5]\text{mV/A}\). Typical trajectory segments are plotted in Fig. 3(e). The computed domain velocities \( v_D \) from AIGLE and from MD are plotted as a function of \( E \) in Fig. 3(d). They show good agreement for \( E \leq E^* \), but, for \( E > E^* \), \( v_D \) from MD becomes significantly
larger than that from GLE, because on the MD time scale the system cannot fully relax and moves farther out of equilibrium as the applied electric field increases. For 
\[ E \leq E^*, \] AIGLE, which was fully parameterized with \textit{ab initio} data, is in good agreement with the experimental Merz’s law \( \log v_D = \log v_0 - E_0/E \). The best fit for \( E \in [1.6, 2.4] \text{mV/A} \) is plotted in the inset of Fig. 3(d), showing excellent agreement with the MD data.

In these comparisons, AIGLE shows striking accuracy in describing rare domain motions microscopically activated by nucleation \([34]\). This is due to a faithful inclusion of memory effects and noise, which can not be achieved if one used instead a Markovian approximation in the Langevin equation. The importance of memory effects can be well illustrated by using the standard Langevin equation in place of AIGLE, with the same external forces and a friction \( \vartheta \) given by
\[ \vartheta = \sum_{s=0}^{n-1} K_{s+\frac{1}{2}} \Delta t, \] i.e., free of memory effects, and a Gaussian white noise \( \xi(n) \) with amplitude determined by 2FDT. As shown in the inset of Fig. 3(d), Markovian Langevin dynamics only agrees with AIGLE and MD for \( E \approx E^* \), but underestimates \( v_D \) by orders of magnitude when the dynamics is glassy.

The regime of small electric fields \( E \leq E^* \) determines the shape of the hysteresis loop in ferroelectric switching phenomena. With the Merz’s law behavior extracted from AIGLE, we compute the hysteresis loop for switching thin films of PbTiO\(_3\) of different radii \( l \). Experimentally, the hysteresis loop depends not only on the switched area \( \pi l^2 \), but also on the shape of the pulse. Here we simply assume square pulses of \( E \) lasting \( t_p \) and approximate the lower branch of the hysteresis loop with a linear model
\[ \ell P(E) = v_D(E) t_p (P_0 + \chi E) - (1 - v_D(E) t_p) (P_0 - \chi E), \] where \( v_D(E) t_p \) is restricted to \([0, l]\). The remnant polarization \( P_0 \) and the susceptibility \( \chi \) are extracted from the MD simulations. The upper branch of the hysteresis loop is computed in a similar way. The results are plotted in Fig. 3(f) and compared to experiments \([70, 73]\). Overall the agreement between theory and experiment is impressive given that the theoretical model did not use any empirical input. Relatively small quantitative discrepancies between theory and experiment may be due to pinning effects of point defects \([75]\) present in the experimental samples. Further refinements of the hysteresis loop calculation will require analysis of the finite thickness effects \([76, 77]\) and consideration on the curvature dependence of the domain wall velocity.

\section*{DISCUSSION}

The examples presented here illustrate how AIGLE is consistently built on atomistic models derived from DFT. AIGLE is capable of modeling rare events on glassy landscapes caused by atomic-scale nucleation-and-growth events. It systematically reproduces MD results within a predictable and explainable near-equilibrium regime \( |E| \leq E^* \), with a much lower computational cost than atomistic simulations. The present work is focused on the near-equilibrium regime in which noise-induced rare events are important. The methodology presented here is also applicable to regimes far from equilibrium. In that case, the effect of the noise might be overwhelmed by the effect of the external driving force, but the memory effect might still be important and the memory term might still be quite complex. This is a direction that we intend to explore in the near future. Another important issue is how to extend the current AIGLE approach by including position dependence in the memory kernel and the noise generator. In general, the CVs are not spatially homogeneous. For example, if one chooses to study the dynamics of the radius of a spherical nucleus associated with a typical first-order phase transition such as the crystallization of water, the memory kernel should always have a weak and probably monotonic dependence on the radius. This type of generalization will also be considered in the future.

The case study presented here is a prelude to \textit{ab initio} multi-scale modeling of more general domain dynamics and phase separation in typical condensed phase systems such as ferromagnets \([78]\), ferroelectrics \([79]\), alloys \([10, 11]\), etc. Although the current work is limited to a uni-variant CV, the generalization to multi-variant CV should be relatively straightforward for CVs with a sparse correlation matrix. Thus, we are looking forward to an \textit{ab initio} discretized field theory built on AIGLE and applied to the problems described above. Indeed, it should be possible to apply this approach to wealth of different problems, for instance, modeling morphology evolution in epitaxial growth \([80, 82]\) or height fluctuation of two dimensional membranes \([83, 84]\), important for the fabrication and the characterization of nanomaterials. In applications to soft matter, AIGLE may provide coarse grained force fields \([80, 85]\) with dynamics compatible with that of all-atom simulations. As a consequence, slow biological processes could be modeled directly without resorting to complex enhanced sampling techniques \([86, 87]\). We leave these studies to future work.

\section*{METHODS}

\subsection*{Training of AIGLE}

In the following, we illustrate the learning procedure with the uni-variant GLE
\[ a(n) = F(n) + \sum_{s=0}^{n-1} K_{s+\frac{1}{2}} v_{n-s-\frac{1}{2}} \Delta t + R(n), \] and the GAR model.
\[ R(n) = \sum_{k=1}^{m_A} \phi(k) R(n-k) + \mu(n) + \sigma(n) w(n), \]

\begin{multline}
\sum_{s=0}^{n-1} K_{s+\frac{1}{2}} v_{n-s-\frac{1}{2}} \Delta t + R(n) \end{multline}
The scalar mass is absorbed in Eq. (7) without loss of generality.

**Separation of noise**

The first step of learning relies on equilibrated MD trajectories $\kappa = \{x(n) | n \in [0, N]\}$ with ergodic fast degrees of freedom. $v_{(n+\frac{1}{2})}$ and $a_n$ are computed from Eq. (3). $v_{(n)}$ is further determined as the average of $v_{(n+\frac{1}{2})}$ and $v_{(n-\frac{1}{2})}$. We can turn the ensemble-averaged orthogonality condition $\langle R_{(n)} v_{(0)} \rangle = 0$ to a time-averaged one. To do so we introduce the shifted GLE with an arbitrary starting point $n_0 \geq 0$:

$$a_{(n)} = F_{(n)} + \sum_{s=0}^{n-n_0-1} K_{(s+\frac{1}{2})} v_{(n-\frac{1}{2}-s)} \Delta t + \tilde{R}_{(n)}^{(n_0)}.$$

$n > n_0$ is required. As demonstrated in Ref. [32], $\langle \tilde{R}_{(n_0+k)}^{(n_0)} \tilde{R}_{(n_0)}^{(n_0)} \rangle = \langle R_{(n+k)} R_{(n)} \rangle$ for a stationary noise series when $n \to \infty$.

For a given CV trajectory, the shifted noise $\tilde{R}_{(n)}^{(n_0)}$ is explicitly computed by inverting Eq. (9):

$$\tilde{R}_{(n)}^{(n_0)} = a_{(n)} - F_{(n)} - \sum_{s=0}^{n-n_0-1} K_{(s+\frac{1}{2})} v_{(n-s-\frac{1}{2})} \Delta t.$$

For $n = n_0$, we let $\tilde{R}_{(n)}^{(n)} = a_{(n)} - F_{(n)}$. The time-averaged estimator of $\langle R_{(k)} v_{(0)} \rangle$ can be written as

$$\zeta_k = \frac{1}{N_k} \sum_{n_0=0}^{N_k-1} \tilde{R}_{(n_0+k)}^{(n_0)} v_{(n_0)},$$

where $N_k = N - k$. Note that $\zeta_0$ only depends on the force fields while $\zeta_k$ also depends on the memory kernel for $k > 0$. It is not recommended, for numerical stability, to train $\mathcal{F}$ by imposing the orthogonality condition $\zeta_0 = 0$ directly. We recommended instead to train $\mathcal{F}$ by minimizing the noise in a maximum-likelihood perspective and further decouple the training of $\mathcal{F}$ and $\mathcal{K}$ for stability and efficiency. To achieve these goals, we first define the constrained optimization problem:

$$\text{minimize}_{F^\theta, K^\theta} \sum_{n=0}^{N-1} \sum_{n=m_K} |\tilde{R}_{(n)}^{(0)}|^2 \text{subject to } \sum_{n_0=0}^{N_k-1} \tilde{R}_{(n_0+k)}^{(n_0)} v_{(n_0)} = 0, k \in [1, m_K].$$

The ensemble average $\mathbb{E}_{\kappa \sim \pi_\kappa}$ is not necessary when $\kappa$ is ergodic and sufficiently long. But in practice averaging over multiple finite-size trajectories is preferred. $m_K$ is the finite memory cutoff of $K$. $F^\theta$ and $K^\theta$ are the parameters of $\mathcal{F}$ and $\mathcal{K}$ respectively. $\mathcal{F}$ can be any differentiable parameterized function, including neural networks. Eq. (12) should be transformed into an unconstrained problem for practical application. Notice that the constraint $\mathbb{E}_{\kappa \sim \pi_\kappa} \zeta_k = 0$ can be written as equivalently

$$\mathbb{E}_{\kappa \sim \pi_\kappa} \sum_{n_0=0}^{N_k-1} (a_{(n_0+k)} - F_{(n_0+k)} v_{(n_0)} = \sum_{s=0}^{k-1} K_{(s+\frac{1}{2})} \Delta t \sum_{n_0=0}^{N_k-1} v_{(n_0+k-s-\frac{1}{2})} v_{(n_0)}).$$

Considering $k \in [1, m_K]$, Eq. (13) can be written in matrix form $\mathcal{Y} = \mathcal{C} \mathcal{K}$. $\mathcal{Y}$ and $\mathcal{K}$ are vectors of length $m_K$. $\mathcal{C}$ is a $m_K \times m_K$ lower triangular matrix. The left hand side of Eq. (13) is the $k$-th entry of $\mathcal{Y}$. The $j$-th entry of $\mathcal{K}$ is $K_{(j-\frac{1}{2})} \Delta t$. And $C_{kj} = \mathbb{E}_{\kappa \sim \pi_\kappa} \sum_{n_0=0}^{N_k-1} v_{(n_0+k-j+\frac{1}{2})} v_{(n_0)}$ when $m_K \geq k \geq j \geq 1$. Hence, the least-square solution to Eq. (13) can be written as $\mathcal{K} = \text{Inv}(\mathcal{C}^T \mathcal{C}) \mathcal{C}^T \mathcal{Y}$. Inv is the pseudo-inverse computed from single-value decomposition with a cutoff ratio for avoiding instability.

We are then able to approach Eq. (12) practically by interleaving $n^{GD} \geq 1$ unconstrained optimization steps towards

$$\text{minimize} \sum_{n=0}^{N-1} |\tilde{R}_{(n)}^{(0)}|^2 \text{with one iteration of} \mathcal{K} \to (1 - \epsilon) \mathcal{K} + \epsilon \text{Inv}(\mathcal{C}^T \mathcal{C}) \mathcal{C}^T \mathcal{Y},$$

The parameter $\epsilon \in (0, 1)$ should be small enough for stability. In this work we use $\epsilon = 0.01$. The second step in Eq. (15) forces $K_{(m_K)} = 0$ over the course of training.

**Training of the GAR model**

In the previous step, the noise $R_{(n)} = \tilde{R}_{(n)}^{(0)}$ is extracted from $a_{(n)}$. Then one can establish a GAR model with $R_{(n)}$ as data. The parameters of the GAR include the linear coefficients $\phi = (\phi_1, \cdots, \phi_k)$ and the parameters $\{\mu^\theta, \sigma^\theta\}$ of the neural network. We define the maximum likelihood loss function

$$L = \mathbb{E}_{\kappa \sim \pi_\kappa} \sum_{n} \log \sigma^2_{(n)} + \left( \frac{R_{(n)} - \sum_{k=1}^{m_A} \phi_{(k)} R_{(n-k)} - \mu_{(n)}}{\sigma^2_{(n)}} \right)^2.$$

It is not recommended to minimize $L$ directly with respect to all the parameters without constraints. Overfitting the data should be avoided for the long-term stationarity of the GAR model. This is crucially important for simulating AIGLE at or above the $\mu$s scale, much longer than the picosecond/nanosecond duration of MD trajectories. So we harness the GAR model
by imposing on $\phi$ the constraint that they should satisfy the Yule-Walker equation. For a given CV trajectory, let $\lambda_{(k)}$ be the estimator of the noise ACF, given as $\lambda_{(k)} = \frac{1}{N_k-m_A} \sum_{n=m_A}^{N_k-1} R(n+k)R(n)$. Let the vector $\Lambda$ be $\Lambda = E_{\kappa \sim p_\kappa}(\lambda(1), \cdots, \lambda(m_A))$. Let the $m_A \times m_A$ matrix $M$ be $M_{jk} = E_{\kappa \sim p_\kappa}(\lambda_{(j-k)})$. The Yule-Walker equation for a standard AR($m$) model is $\Lambda = \mathcal{M} \phi$, the least square solution of which can be written as $\phi_{YW} = \text{Inv}(\mathcal{M}^T \mathcal{M}) \mathcal{M}^T \Lambda$. Using the Yule-Walker solution as a constraint, we optimize the GAR model by interleaving $n^{SD}$ unconstrained optimization steps towards

$$
\text{minimize } L \quad (17)
$$

with one iteration of

$$
\phi \rightarrow (1 - \epsilon) \phi + \epsilon \text{Inv}(\mathcal{M}^T \mathcal{M}) \mathcal{M}^T \Lambda. \quad (18)
$$

Although in the formal presentation the training of GAR is done after the training of the first step, in practice one can train GAR on the fly to simplify the implementation.

**Incorporation of near-equilibrium data**

In this step, we deal with additional datasets that violate detailed balance. We fix the memory kernel and the GAR model obtained for thermal equilibrium, assuming that they are approximately the same in near-equilibrium situations. The optimization task is simply

$$
\text{minimize } E_{F^\theta} \sum_{\kappa \sim p_\kappa} \sum_{n=m_K}^{N-1} |\tilde{R}_n(0)|^2 \quad (19)
$$

for the extended dataset. Here $F^\theta$ may include the parameters for the external driving forces.

**Details of the harmonic chain**

Let $\{z_i\}_{i \in [0,N]}$ and $\{p_i\}_{i \in [0,N]}$ be the coordinate and the momentum of the particles in the harmonic chain. Let $z_0 = 0$ and $p_0 = 0$. The Hamiltonian of the system is

$$
H(\{p_i\}, \{z_i\}) = \sum_{i=0}^{N-1} \frac{p_i^2}{2m} + \frac{1}{2} \omega^2 (z_i - z_{i+1} - a)^2. \quad (20)
$$

The parameters of the system we simulate are $N = 100$, $m = k_BT = \omega = 1$, $x_N$ is connected to a Langevin thermostat with damping time $t_d = 10$. The CV is $x = x_N$. We use a MD time step $\delta t = 0.1$ and obtain 200 trajectories of $x$, each lasting $t_{MD} = 2000$ after initial equilibration. The trajectories are further coarse-grained with a time step $\Delta t = 4 \delta t$ by sampling the MD data every 4 MD time steps. The GLE ansatz for this system can be written as

$$
a_n = -\omega_0^2 x_n + \sum_{s=0}^{n-1} K(s + \frac{1}{2}) \varphi(n-s - \frac{1}{2}) \Delta t + R(n). \quad (21)
$$

The neural network part of the corresponding GAR model is a feed-forward neural network with two hidden layers (size=10). The neural network only outputs $\mu(n)$ in Eq. (5). The $\sigma(n)$ in Eq. (5) is parameterized directly by one scalar variable since its history dependence is unnecessary for this case study. The finite memory cutoff is taken to be $m_K = 50$ for the memory kernel and $m_A = 25$ for the GAR model. The training is done by the Adam optimizer implemented in PyTorch [88] with the default setting and learning rate of 0.001. Convergence is reached with approximately 3000 iterations. For each iteration, 50 trajectories of the CV are randomly chosen for optimization. The parameter $\epsilon$ for both Eq. (15) and Eq. (18) is 0.01. The interval $n^{SD}$ is 10. The optimized $\omega_0$ is 0.09. The optimized standard deviation of the white noise is $\sigma(n) = 0.721$. The colored noise $R$ and the white noise $\omega$ are both distributed normally on the dataset with negligible off-centering. Non-stationarity is not detected numerically for the GAR model.

**Details of the ab initio domain motion modeling**

**The atomistic models**

The atomistic modeling of PbTiO$_3$ is based on the methodology developed in Refs. [19, 30, 91]. The DP model for representing the potential energy surface is trained on DFT data with the SCAN meta-GGA functional [68]. The dataset is collected through the active learning procedure implemented in the DPGEN code [92, 83]. The data are labelled by Quantum ESPRESSO [93] with norm-conserving pseudo-potentials [97] including semi-core states. An early version of the current DP model was introduced in Ref. [91]. It was trained with configurations without domain walls and used to study the ferroelectric phase transition in bulk PbTiO$_3$ with atomistic simulations. The model predicted thermodynamic and ferroelectric properties of PbTiO$_3$ in close agreement with experimental results, and provided unique insight on the microscopic mechanism driving the phase transition. The numerical protocol we follow to train the current DP model is identical to Ref. [91], which contains the full technical details. The final dataset for the current DP model contains configurations without domain walls and with two twin $90^\circ$ domain walls or two twin $180^\circ$ domain walls. 1276 configurations with no domain walls are collected with the same active learning procedure of Ref. [91] in the temperature interval $[300K, 1200K]$ and pressure interval $[0, 10^6 Pa]$. 494 configurations with domain walls are collected with the same active learning procedure in the temperature interval $[300K, 1200K]$ and pressure interval $[0, 10^6 Pa]$.
The blue color marks configurations without domain walls, and the orange color marks the rest. $E_{\text{DFT}}, F_{\text{DFT}}^\alpha$ are the energy and force labels of the data. $E_0$ is a constant. $\Delta E, \Delta F_{x,y,z}$ are the difference between the model prediction and the data label.

[100K, 600K] and the pressure interval [0, 10^5Pa]. The largest supercell with domain wall configurations contains 54 elementary cells, corresponding to 270 atoms. We use a spatial cutoff radius of 8Å for the DP model. This is larger than in typical DP models but is necessary to capture longer range interactions in th domain wall region.

Fig. 4 plots the nearly gaussian error distribution of the DP model against the training set. The root-mean-square error (RMSE) of energy is 1meV/atom for configurations without domain walls and 1.4meV/atom for the remaining ones. The RMSE of the Cartesian components of the forces is around 0.1eV/Å for all cases. The force error is slightly larger for the direction orthogonal to the domain walls, which is the x-axis for our data with 180° domain walls. The DP model is also validated on a small independent testing set, showing similar accuracy to its performance on the training set. We conclude that the DP model reproduces statistically the adiabatic potential energy surface of SCAN-DFT.

Next, we calculate the Berry-phase polarization change, relative to the centrosymmetric structure for all configurations without domain walls. The calculation is done with the Wannier90 code [96]. The total polarization is the sum of electronic (Berry phase) contributions and ionic contributions [97]. The dipole $D$ of the simulation cell (cell dipole) is obtained by multiplying the total polarization by the volume of the cell. Born charge (BC) associated to atom- $i$ is a tensor $Z_{i,\alpha\beta} = \frac{\partial D_{\alpha}}{\partial x_i}$ defined as the change of $D$ in direction $\alpha$ caused by the displacement of atom- $i$ away from its equilibrium position in direction $\beta$. With BC, a linear approximation to $D_\alpha$ can be written as $D_\alpha = \sum_i \sum_{\beta=x,y,z} Z_{i,\alpha\beta} \delta x_i, \beta$, referred to as the BC model. Due to space-group symmetry or negligible contribution, some degrees of freedom of BC can be eliminated. For PbTiO$_3$, the BC of Pb and Ti atom is conventionally approximated [98] by diagonal matrix $Z_{\text{Pb,} \alpha\beta} = Z_{\text{Pb},\alpha\beta}$ and $Z_{\text{Ti,} \alpha\beta} = Z_{\text{Ti},\alpha\beta}$, respectively. The BC of O atom is also approximated diagonally, with value $Z_{\text{O1}}$ for Ti-O bond direction, and $Z_{\text{O2}}$ for the other two orthogonal directions. In the present study, the BC model is fitted to the cell dipole data with mean squared error loss, leading to $Z_{\text{Pb}} = 3.7140e$, $Z_{\text{Ti}} = 5.4897e$, $Z_{\text{O1}} = -3.3551e$ and $Z_{\text{O2}} = -2.9234e$. The RMSE of the cell dipole predicted by the resultant BC model is 2eÅ for a 3 x 3 x 3 supercell. This corresponds to roughly a 2µC/cm$^2$ RMSE in polarization, much smaller than the 72µC/cm$^2$ total polarization of PbTiO$_3$ extracted from our room temperature simulations. The BC model constructed in this way is statistically more accurate at finite temperature than models using conventional Born effective charges calculated by perturbing the equilibrium structures [99]. Using the BC model defined above, the local electric dipole $p_j$ associated with each Ti-centered elementary cell- $j$ of PbTiO$_3$ is given by a weighted sum of the position of atoms in the cell- $j$ [60, 61]. The local dipole defined via the BC model is an approximation of the local dipole introduced in Ref. [91] via maximally localized Wannier functions. The latter includes full (nonlinear) environmental dependence of local dipoles, computationally much more expansive than the simple BC model. So in the present study, we use BC model instead, for accelerating the simulation of domain wall dynamics. In addition, the linear approximation is sufficiently accurate for MD simulations at room temperature, which is far away from the ferroelectric phase transition temperature of PbTiO$_3$.

The MD simulations

With the BC model, electric-field driven (non equilibrium) MD can be simulated from the evolution under a Hamiltonian $H$ obtained by adding interaction of the dipoles with the external field $E$ to the Hamiltonian $H_0$ in absence of the field, i.e., $H = H_0 - \sum_j E \cdot p_j$. We use LAMMPS [55] and PLUMED [100] for these simulations.

In our simulations in the NP, T ensemble, we approximate the environmental noise by applying a Langevin thermostat to the thin atomic layers highlighted in pink in Fig. 2 which contain atoms far away from the initial locations of the domain walls. The cell tensor is diagonal and kept fixed along the x, y direction (lattice constant $a = b = 3.91 Å$) to mimic the experimental lattice constants in the plane imposed by the substrate in the epitaxially grown thin films. A barostat is applied in the
The training of AIGLE

The MD trajectory of the CV $x$ is coarse-grained through the application of truncated gaussian filters:

$$x(n) = \frac{1}{A_{\text{wall}}} \sum_{q=0}^{3\sigma} \sum_{s=0}^{3\sigma} \exp(-\frac{x^2}{2\sigma^2}) \alpha(t = n\Delta t - q\delta t).$$

(22)

where $\Delta t = 10fs$ and $\delta t = 2fs$. The constant $A_{\text{wall}} = 400$ is the area of the supercell in the $xz$ plane. The truncation parameter $\sigma = 40$, corresponding to a timescale of 0.1ps, is sufficient to filter out the high-frequency vibrational modes of the local dipole. The effective mass of the CV $x$ estimated from the equipartition theorem is $m = 1.0 \times 10^{3}u$. The GLE ansatz for $x$ can be written as

$$a(n) = -\partial_{x}U(x(n)) + pE + \sum_{s=0}^{n-1} K_{(s+\frac{1}{2})} \partial_{x}(a_{(s+\frac{1}{2})}) \partial_{x} \Delta t + \frac{1}{m} R(n).$$

(23)

The cutoff of the memory kernel $K$ and the GAR model is $m_K = 200$ and $m_A = 40$ respectively.

In the first two training steps of the AIGLE model, the dataset consists of 400ps long trajectories of $x$ under $E = 2mV/A$. The underlying Hamiltonian dynamics is at metastable equilibrium — the twin domain walls are trapped by the periodic potential around a local minimum within the simulated time scale. Under such circumstance, the potential energy surface of CV is not fully explored but the fast atomistic degrees of freedom are sufficiently ergodic for determining $K$ and the GAR model. The GAR model contains a feed-forward neural network with two hidden layers (size=10). The $\sigma(n)$ is parameterized by one scalar variable. We use the Adam optimizer with an initial learning rate of 0.01 and an exponential decay rate of 0.9 every 500 steps. We let $\eta^{GD} = 10$ and $\epsilon = 0.01$. Convergence is reached within 5000 iterations. The entire dataset is used for every iteration. After the training, the colored noise $R$ and the white noise $\omega$ are both distributed normally on the dataset with negligible off-centering. Non-stationarity is not detected numerically for the GAR model.

In the out-of-equilibrium regime, we fix the memory kernel and the GAR model. Then we retrain the force field $F(x) = -\partial_{x}U(x) + pE$ with MD trajectories (about 1.2ns long in total) simulated independently under different $E \in [2.0, 2.4]$mV/A. In this dataset, the MD can be out-of-equilibrium by creeping down metastable states. Although the entire landscape is explored, the creeping events are too sparse in the dataset for fitting the barrier height $U_b$ accurately. So we predetermine it using metadynamics, a method for computing free energy differences [34], under $E = 0$. The resulting free energy barrier $\Delta E$ in energy unit is converted to $U_b = \Delta E/\epsilon^2$, where $\epsilon = 3.91A$ is the lattice constant. All the other parameters in $F(x)$ are then trained directly on the MD dataset. We use the Adam optimizer with the same setting. Convergence is reached within 5000 iterations. The entire dataset is used for every iteration.

ACKNOWLEDGEMENT

We thank Yucheng Yang, William M Jacobs, Yixiao Chen, and Linfeng Zhang for fruitful discussions. All authors were supported by the Computational Chemical Sciences Center: Chemistry in Solution and at Interfaces (CSI) funded by DOE Award DE-SC0019394 and the Princeton Research Computing resources at Princeton University which is consortium of groups led by the Princeton Institute for Computational Science and Engineering (PICSciE) and Office of Information Technology’s Research Computing. P.X. and R.C. were supported by the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231. P.X. and W.E were also supported by a gift from iFlytek to Princeton University.

[1] N. Provatas and K. Elder, Phase-field methods in materials science and engineering (John Wiley & Sons, 2011).
[2] F. Müller-Plathe, ChemPhysChem 3, 754 (2002).
[3] M. G. Saunders and G. A. Voth, Annual review of biophysics 42, 73 (2013).
[4] E. Brini, E. A. Algaer, P. Ganguly, C. Li, F. Rodríguez-Ropero, and N. F. van der Vegt, Soft Matter 9, 2108 (2013).
[5] S. Kmiecik, D. Gront, M. Kolinski, L. Wieteska, A. E. Dawid, and A. Kolinski, Chemical Reviews 116, 7898 (2016).
[66] S. Liu, I. Grinberg, and A. M. Rappe, Nature 534, 360 (2016).
[67] W. J. Merz, Phys. Rev. 95, 690 (1954).
[68] J. Sun, A. Ružičsinszky, and J. P. Perdew, Phys. Rev. Lett. 115, 036402 (2015).
[69] T. Morita and Y. Cho, Jpn. J. Appl. Phys. 43, 6535 (2004).
[70] W.-H. Kim, S. M. Yoon, and J. Y. Son, Mater. Lett. 124, 47 (2014).
[71] R. Nishino, T. C. Fujita, F. Kagawa, and M. Kawasaki, Scientific reports 10, 1 (2020).
[72] Ø. Dahl, J. K. Grepstad, and T. Tybell, J. Appl. Phys. 106, 084104 (2009).
[73] T. Morita and Y. Cho, Jpn. J. Appl. Phys. 43, 6535 (2004).
[74] L. Cheng and N. K. Yip, Phys. D: Nonlinear Phenom. 297, 1 (2015).
[75] L. He and D. Vanderbilt, Phys. Rev. B 68, 134103 (2003).
[76] V. Nagarajan, I. Jenkins, S. Alpay, H. Li, S. Aggarwal, L. Salamanca-Riba, A. Roytburd, and R. Ramesh, J. Appl. Phys. 86, 595 (1999).
[77] M. Dawber, P. Chandra, P. Littlewood, and J. Scott, J. Phys. Condens. Matter. 15, L393 (2003).
[78] N. Faghihi, N. Provatas, K. Elder, M. Grant, and M. Karttunen, Phys. Rev. E 88, 032407 (2013).
[79] L.-Q. Chen, J. Am. Ceram. Soc. 91, 1835 (2008).
[80] B. Wu, D. Geng, Z. Xu, Y. Guo, L. Huang, Y. Xue, J. Chen, G. Yu, and Y. Liu, NPG Asia Mater. 5, e36 (2013).
[81] J. Zhuang, W. Zhao, L. Qiu, J. Xin, J. Dong, and F. Ding, J. Phys. Chem. C 123, 9902 (2019).
[82] K. Momeni, Y. Ji, K. Zhang, J. A. Robinson, and L.-Q. Chen, NPJ 2D Mater. Appl. 2, 1 (2018).
[83] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al., Advances in neural information processing systems 32 (2019).
[84] L. Zhang, J. Han, H. Wang, W. Saidi, R. Car, and W. E, Advances in Neural Information Processing Systems 31 (2018).
[85] L. Zhang, M. Chen, X. Wu, H. Wang, W. E, and R. Car, Phys. Rev. B 102, 041121 (2020).
[86] Y. Zhang, H. Wang, W. Chen, J. Zeng, L. Zhang, H. Wang, and W. E, Comput. Phys. Commun. 253, 107206 (2020).
[87] R. Resta and D. Vanderbilt, “Theory of polarization: A modern approach," in Physics of Ferroelectrics: A Modern Perspective (Springer Berlin Heidelberg, Berlin, Heidelberg, 2007) pp. 31–68.
[88] U. V. Waghmare and K. M. Rabe, Phys. Rev. B 55, 6161 (1997).
[89] J. Zhuang, W. Zhao, L. Qiu, J. Xin, J. Dong, and F. Ding, J. Phys. Chem. C 123, 9902 (2019).
[90] K. Momeni, Y. Ji, K. Zhang, J. A. Robinson, and L.-Q. Chen, NPJ 2D Mater. Appl. 2, 1 (2018).
[91] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al., Advances in neural information processing systems 32 (2019).
[92] L. Zhang, J. Han, H. Wang, W. Saidi, R. Car, and W. E, Advances in Neural Information Processing Systems 31 (2018).
[93] L. Zhang, M. Chen, X. Wu, H. Wang, W. E, and R. Car, Phys. Rev. B 102, 041121 (2020).
[94] P. Xie, Y. Chen, W. E, and R. Car, arXiv preprint arXiv:2205.11839 (2022).
[95] L. Zhang, D.-Y. Lin, H. Wang, R. Car, and W. E, Phys. Rev. Materials 3, 023804 (2019).
[96] L. Zhang, H. Wang, W. Chen, J. Zeng, L. Zhang, H. Wang, and W. E, (2019).
[97] R. Resta and D. Vanderbilt, “Theory of polarization: A modern approach," in Physics of Ferroelectrics: A Modern Perspective (Springer Berlin Heidelberg, Berlin, Heidelberg, 2007) pp. 31–68.
[98] U. V. Waghmare and K. M. Rabe, Phys. Rev. B 55, 6161 (1997).
[99] W. Zhong, D. Vanderbilt, and K. M. Rabe, Phys. Rev. B 52, 6301 (1995).
[100] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi, Comput. Phys. Commun. 185, 604 (2014).