



**HAL**  
open science

# Modeling uncertainties in molecular dynamics simulations using a stochastic reduced-order basis

Haoran Wang, Johann Guilleminot, Christian Soize

► **To cite this version:**

Haoran Wang, Johann Guilleminot, Christian Soize. Modeling uncertainties in molecular dynamics simulations using a stochastic reduced-order basis. *Computer Methods in Applied Mechanics and Engineering*, 2019, 354, pp.37-55. 10.1016/j.cma.2019.05.020 . hal-02146341

**HAL Id: hal-02146341**

**<https://hal.science/hal-02146341>**

Submitted on 3 Jun 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Modeling uncertainties in molecular dynamics simulations using a stochastic reduced-order basis

Haoran Wang<sup>a</sup>, Johann Guilleminot<sup>a,\*</sup>, Christian Soize<sup>b</sup>

<sup>a</sup>Department of Civil and Environmental Engineering, Duke University, Durham, NC 27708, USA

<sup>b</sup>Laboratoire Modélisation et Simulation Multi Echelle MSME, UMR 8208 CNRS, Université Paris-Est, 5 bd Descartes, 77454 Marne-la-Vallée, France

---

## Abstract

A methodology enabling the robust treatment of model-form uncertainties in molecular dynamics simulations is proposed. The approach consists in properly randomizing a reduced-order basis, obtained by the method of snapshots in the configuration space. A multi-step strategy to identify the hyperparameters in the stochastic reduced-order basis is further introduced. The relevance of the framework is finally demonstrated by characterizing various types of modeling errors associated with molecular dynamics simulations on a graphene sheet. In particular, the ability of the framework to represent uncertainties raised by model reduction and interatomic potential selection is assessed.

*Keywords:* Model Uncertainties, Molecular Dynamics, Stochastic Reduced-Order Modeling, Uncertainty Quantification

---

## 1. Introduction

Molecular Dynamics (MD) simulations are widely used in computational chemistry, materials science, and multiscale approaches. Applications include the prediction of material behavior under normal and extreme conditions, the interpretation of small-scale experimental observations and the design of new functionalized materials, to list a few. The construction and identification of relevant functional forms for interatomic potentials is a central aspect of all MD simulations. These intricate issues are usually undertaken with the aim to accurately estimate some problem-specific quantities of interest. As a result, several modeling solutions can coexist for a given system of (potentially coarse-grained) particles and can lead to very different predictions at the macroscopic level. Graphene is a relevant example for which many potentials were proposed, including the Morse, Tersoff, Airebo, and LCBOP potentials (see [1], [2], [3], and [4] respectively). In addition, identified potential parameters often vary from one study to another, and nine versions of Tersoff-style potentials can be found for graphene [5, 2, 6, 7, 8, 9, 10, 11, 12]. While the issue of selecting interatomic potentials can be in part circumvented by resorting to *ab-initio* MD simulations (for reasonably small systems), it remains a major bottleneck that can drastically affect the fidelity of the simulations.

In recent years, there has been a growing interest in the quantification and propagation of parametric uncertainties in MD simulations [13, 14], atomistic-to-continuum coupling strategies (see, *e.g.*, [15, 16]), and calibration and validation strategies in the presence of uncertainties (see [17, 18] and the references therein); see [19] for a general review about uncertainty quantification techniques. For instance, the effect of a stochastic Lennard Jones potential depth on the ionic conductance of Na<sup>+</sup> and Li<sup>+</sup> through a nanopore in SiO<sub>2</sub> was studied in [20], and the Bayesian solving of an inverse problem for the stochastic potential parameters of water molecules, based on density, enthalpy, and self-diffusion, was performed in [21]. The case of model-form uncertainties, generated by the *a priori* selection of functional forms for interatomic potentials, has received very little attention to date. A methodology to estimate the sensitivity of MD predictions with respect to empiric potentials was proposed in [22]. The proposed strategy provides a path towards the identification of relevant sources of modeling errors and allows errors in thermodynamical averages to be corrected between two models.

The aim of this work is to provide a probabilistic framework to assess, model and propagate model-form uncertainties in MD simulations. The approach is based on the randomization of a reduced-order basis, using a random matrix approach [23]. The stochastic reduced-order model thus constructed is specifically introduced as a mean to account for modeling errors in a nonparametric way, acting on the coefficients of the dynamics in the low-dimensional space. In contrast to the approach advanced in [22], the methodology does not rely on an infinitesimal perturbation around a given potential, but rather consists of constructing a probabilistic model that is identified on a family of potentials that represent different functional forms and, possibly, different numerical parameterizations of the same potential. This strategy enables a complete probabilistic description and propagation of uncertainties, generating a family of random systems that includes random-system realizations that can be “far” from a nominal system, defined through a reference potential.

This paper is organized as follows. The molecular dynamics background is briefly reminded in Section 2. The construction of the stochastic reduced-order basis is then presented in Section 3. Applications are provided in Section 4.

### List of notations.

- $d$ : dimension of the physical domain.
- $n$ : reduced-order dimension.
- $\mathbf{f}$ : global force vector (physical space).
- $\mathbf{f}^i$ : force vector acting on atom  $i$  (physical space).
- $\mathbf{q}$ : global coordinate vector (physical space).
- $\mathbf{q}^i$ : position vector of atom  $i$  (physical space).
- $N$ : total number of dofs before applying zero Dirichlet conditions.
- $N_a$ : total number of atoms in the physical space (standard MD).
- $N_{\text{CD}}$ : total number of dofs subjected to zero Dirichlet conditions.
- $N_\epsilon$ : number of increments of engineering strain.
- $N_s$ : number of snapshots used to construct the reduced-order basis.
- $\mathbb{M}_{N,n}$ : set of all the  $N \times n$  real matrices.
- $\mathbb{M}_n$ : set of all the  $n \times n$  real matrices.
- $\mathbb{M}_n^+$ : set of all the symmetric positive-definite (SPD)  $n \times n$  matrices.
- $\mathbb{M}_n^{+0}$ : set of all the symmetric positive-semidefinite  $n \times n$  matrices.
- $\mathbb{M}_n^u$ : set of all the  $n \times n$  upper triangular real matrices.
- $\mathbb{M}_n^{+u}$ : set of all the  $n \times n$  upper triangular real matrices with strictly positive diagonal.
- $[I_n]$ :  $n \times n$  identity matrix.
- $[0_n]$ :  $n \times n$  zeros matrix.
- $[W]$ : deterministic reduced-order basis (ROB).
- $[\mathbf{W}]$ : stochastic reduced-order basis (SROB).
- $E$ : mathematical expectation.
- Tr: trace operator.
- Var: variance.

## 2. Background in molecular dynamics

### 2.1. Basic equations

Consider a system composed of  $N_a$  atoms, interacting in an open subset  $\Omega \subset \mathbb{R}^d$  over a finite simulation time  $T$ . The dynamics of atom  $i$  is described by the Newton law

$$\begin{cases} m_i \ddot{\mathbf{q}}^i(t) = \mathbf{f}^i(\mathbf{q}^i(t), t), \\ \mathbf{q}^i(0) = \mathbf{q}_0^i, \quad \dot{\mathbf{q}}^{(i)}(0) = \dot{\mathbf{q}}_0^i, \end{cases} \quad (1)$$

where  $m_i$  is the mass of the atom,  $\ddot{\mathbf{q}}^i$  is the acceleration vector corresponding to its position vector  $\mathbf{q}^i$ ,  $\mathbf{f}^i$  is the force exerted on it, and  $(\mathbf{q}_0^i, \dot{\mathbf{q}}_0^i)$  are given initial conditions in  $\mathbb{R}^d \times \mathbb{R}^d$ . The force vector can be written as

$$\mathbf{f}^i(\mathbf{q}^i(t), t) = \mathbf{f}_{\text{int}}^i(\mathbf{q}^i(t)) + \mathbf{f}_{\text{ext}}^i(t), \quad (2)$$

where  $\mathbf{f}_{\text{int}}^i$  and  $\mathbf{f}_{\text{ext}}^i$  represent interatomic contributions from surrounding atoms and external forces specified to prescribe, *e.g.*, boundary conditions. Let  $\mathbf{q}$  and  $\mathbf{f}$  be the vectors of positions and forces for all atoms, gathered component-wise, and denote by  $[M]$  the associated  $N \times N$  diagonal mass matrix (with  $N = d \times N_a$ ). The internal force is derived from a given interatomic potential  $V$ , defined on the configuration space, such that

$$\mathbf{f}_{\text{int}}^i = -\nabla_{\mathbf{q}^i} V. \quad (3)$$

The potential  $V$  typically incorporates several types of interactions, such as bonded and/or long-range interactions. The state of the system can be described in phase space, using the variables  $\mathbf{q}$  and  $\mathbf{p}$ , where  $\mathbf{p}$  denotes the vector of momenta,  $\mathbf{p} = [M]\dot{\mathbf{q}}$ . The dynamics of the whole system can be expressed in a global matrix form as

$$\begin{cases} [M]\ddot{\mathbf{q}}(t) = \mathbf{f}(\mathbf{q}(t), t), \\ \mathbf{q}(0) = \mathbf{q}_0, \quad \dot{\mathbf{q}}(0) = \dot{\mathbf{q}}_0. \end{cases} \quad (4)$$

To simplify the presentation, the action of any control algorithm devised to impose target macroscopic properties (such as temperature, pressure, volume or energy) is omitted.

For later use, we introduce zero Dirichlet boundary conditions acting on a group of  $N_{\text{CD}}$  atoms in the following matrix form:

$$[B]^T(\mathbf{q}(t) - \mathbf{q}(0)) = \mathbf{0}_{N_{\text{CD}}}, \quad t \in [0, T], \quad (5)$$

where  $[B]$  is an  $N \times N_{\text{CD}}$  matrix satisfying, by construction, the orthogonality property  $[B]^T[B] = [I_{N_{\text{CD}}}]$ . In this work, MD simulations are conducted in the NVE ensemble, by using LAMMPS [24] for both standard and reduced-order MD simulations (note that the proposed SROM was implemented in LAMMPS).

## 2.2. Exploring the effect of potential selection: case of a graphene sheet

In order to qualitatively illustrate the impact of potential selection on the MD predictions, we consider the case of a  $40.52 \text{ \AA} \times 15.60 \text{ \AA}$  graphene sheet containing 272 carbon atoms, fixed on its left boundary and subjected to a sinusoidal load  $\mathbf{f}_{\text{ext}}^i(t) = A \sin(2\pi\omega t)\mathbf{e}^2$  for any atom  $i$  on its right boundary [25], with  $A = 6 \text{ kcal.mol}^{-1}.\text{\AA}^{-1}$  and  $\omega = 20 \text{ GHz}$ ; see Fig. 1.

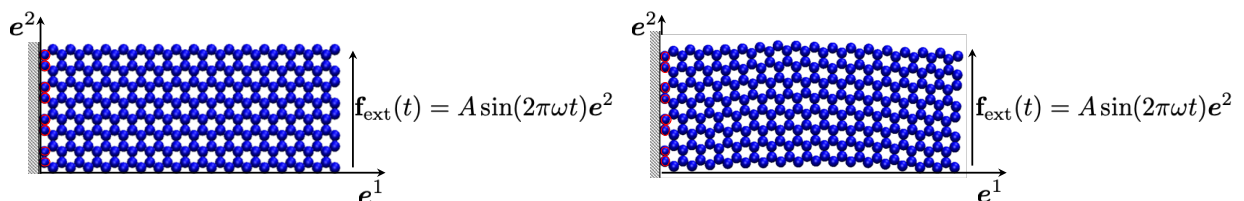


Figure 1: Graphene sheet initial configurations: equilibrium stress-free configuration (left panel) and non-equilibrium configuration (right panel). Red-circled atoms on the left edge satisfy a zero displacement condition, while atoms on the right edge are subjected to force  $f$ .

Figs. 2 and 3 compare the  $\mathbf{e}^1$ - and  $\mathbf{e}^2$ -displacements of all atoms after 60,000 femiseconds (fs). These results are obtained by considering the two different initial configurations shown in Fig. 1 with different interatomic potentials. The impact of potential selection is clearly observed and turns out to be significant. In the case of an equilibrium initial configuration, the maximum magnitude of the  $\mathbf{e}^1$ -displacement for the atoms on the right end, predicted by the modified Morse potential, is twice as much as the one obtained with Tersoff potential. These differences are even more pronounced when the simulation is initialized with a non-equilibrium configuration. In this case, atoms are seen to move to completely opposite directions under the same loading conditions, depending on the interatomic potentials. Such differences at the atomistic

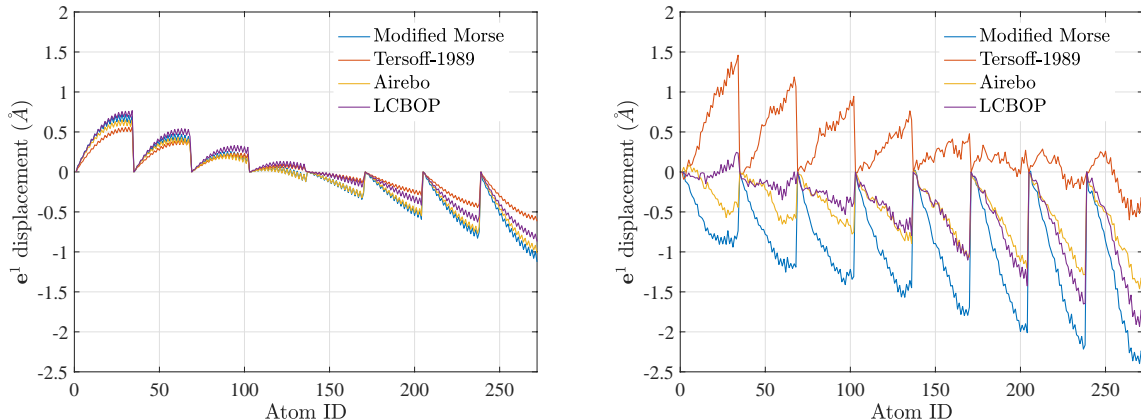


Figure 2: Graph of the  $e^1$ -displacement for all atoms at time  $t = 60,000$  (fs), starting from the equilibrium stress-free configuration (left panel) and the non-equilibrium configuration (right panel) shown in Fig. 1. The results associated with four potentials are displayed: Modified Morse (blue), Tersoff (red), Airebo (orange), and LCBOP (purple). Atoms are numbered from bottom to top and left to right.

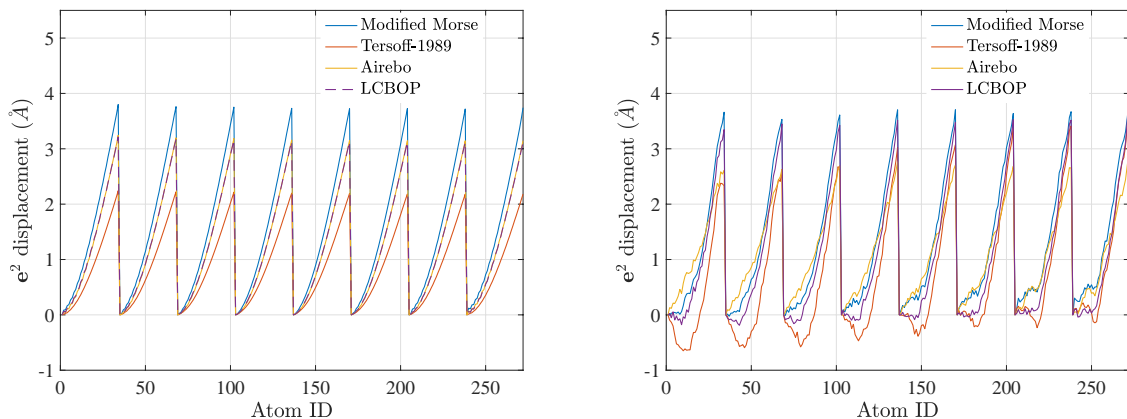


Figure 3: Graph of the  $e^2$ -displacement for all atoms at time  $t = 60,000$  (fs), starting from the equilibrium stress-free configuration (left panel) and the non-equilibrium configuration (right panel) shown in Fig. 1. The results associated with four potentials are displayed: Modified Morse (blue), Tersoff (red), Airebo (orange) and LCBOP (purple). Atoms are numbered from bottom to top and left to right.

scale can give rise to very different continuum characteristics of graphene. Indeed, it has been reported that the elastic moduli can differ by several hundred gigapascals when different potentials are used [26]. In the next section, we present a methodology that allows such model-type uncertainties to be incorporated in MD simulations. The model obtained with the full MD description or any reference model is referred to as the high-fidelity model (HFM) hereinafter.

### 3. Stochastic reduced-order modeling in MD simulations

This section is concerned with the construction of a stochastic reduced-order model for MD computations. To that end, a two-step approach is pursued. A reduced-order model is first introduced in Section 3.1. In a second step, the projection basis is subsequently randomized to account for uncertainties in a nonparametric approach, in Section 3.2. It should be noticed that the reduced-order model is not introduced to save

computation time (which is not efficient here since evaluating the vector of forces necessitates mapping quantities back-and-forth), but rather to enable the proper randomization.

### 3.1. Reduced-order model for molecular dynamics

In this work, a reduced-order model (ROM) is constructed by using a proper orthogonal decomposition (POD) and the method of snapshots [27, 28]. Let  $\{V_\ell, \ell = 1 \dots N_V\}$  be a set of candidates for the interatomic potential (see Eq. (3)). The dependence of the global force vector on the selected potential  $V_\ell$  is denoted by  $\mathbf{f}(\mathbf{q}(t), t; V_\ell)$ . In order to construct the reduced-order basis, for each  $1 \leq \ell \leq N_V$ , we consider the state variable  $\mathbf{z}^\ell(t) \in \mathbb{R}^N$  defined as

$$\mathbf{z}^\ell(t) = \mathbf{q}(t; V_\ell) - \mathbf{q}(0; V_\ell), \quad \forall t \in [0, T]. \quad (6)$$

The function  $\mathbf{z}^\ell$  satisfies

$$[M]\dot{\mathbf{z}}^\ell(t) = \tilde{\mathbf{f}}(t; V_\ell), \quad (7)$$

supplemented with appropriate boundary conditions, with  $\tilde{\mathbf{f}}(t; V_\ell) = \mathbf{f}(\mathbf{z}^\ell(t) + \mathbf{q}(0; V_\ell), t; V_\ell)$ . In this work, two reduced-order bases are constructed: one basis is associated with a given potential  $V_\ell$ , while the other basis is aimed at representing the aforementioned set of potentials. These bases are denoted as  $\{[W_\ell], 0 \leq \ell \leq N_V\}$ , where each element in  $\{[W_\ell], 1 \leq \ell \leq N_V\}$  is the basis associated with each potential in the set  $\{V_\ell, 1 \leq \ell \leq N_V\}$ , and  $[W_0]$  is the basis adapted to all potentials.

For  $\ell \geq 1$ , consider the discretization  $t_0 = 0 < t_1 < \dots < t_{N_t} = T$  of time interval  $[0, T]$  (taken independent of the potential), where  $t_j = j\Delta t$  and  $\Delta t$  is the constant time step. Let  $\mathcal{J} = \{j_1, \dots, j_{N_s(\ell)}\} \subset \{1, \dots, N_t\}$ , with  $1 \leq N_s(\ell) \leq N_t$ , not necessarily ordered and with distinct elements, and consider the sequence of MD snapshots  $\{\underline{\boldsymbol{\mu}}_\ell^{(k)}\}_{k=1}^{N_s(\ell)}$  such that  $\underline{\boldsymbol{\mu}}_\ell^{(k)} = \mathbf{q}(t_{j_k}; \ell)$ .

For  $\ell \geq 1$ , let  $[X_\ell]$  be the  $N \times N_s(\ell)$  matrix defined as

$$[X_\ell] = [\underline{\boldsymbol{\mu}}_\ell^{(1)} \dots \underline{\boldsymbol{\mu}}_\ell^{(N_s(\ell))}], \quad (8)$$

where  $\underline{\boldsymbol{\mu}}_\ell^{(k)} = \boldsymbol{\mu}_\ell^{(k)} - \mathbf{q}(0)$  for  $1 \leq k \leq N_s(\ell)$ . Upon introducing the singular value decomposition

$$[X_\ell] = [U_\ell][S_\ell][V_\ell]^\top, \quad (9)$$

where the singular values are assumed decreasingly ordered, the reduced basis  $[W_\ell]$  is classically obtained by retaining the  $n_\ell$  first columns (which are referred to as POD modes below) of  $[U_\ell]$ . The number  $n_\ell$  of modes can be determined through a convergence analysis enabling a tradeoff between dimensionality reduction and projection error (see Section 4 for an example).

Similarly, let  $[X_0] = [[X_1] \dots [X_{N_V}]]$ , where the number of snapshots  $N_s(\ell)$  is such that convergence is reached for all  $1 \leq \ell \leq N_V$ . Consider next the singular value decomposition (with decreasingly ordered singular values)

$$[X_0] = [U_0][S_0][V_0]^\top. \quad (10)$$

The reduced-order basis  $[W_0]$  is then constructed by retaining the  $n_0$  first columns of  $[U_0]$ .

For  $0 \leq \ell \leq N_V$ , we now introduce the linear mapping

$$\mathbf{z}^\ell(t) = [W_\ell]\mathbf{y}^\ell(t), \quad (11)$$

where  $\mathbf{y}^\ell$  is the reduced variable with values in  $\mathbb{R}^{n_\ell}$ ,  $n_\ell \ll N$ . The matrix  $[W_\ell]$  satisfies the orthogonality property

$$[W_\ell]^\top [W_\ell] = [I_{n_\ell}] \quad (12)$$

and the boundary condition

$$[B]^\top [W_\ell] = [0_{N_{\text{CD}}, n_\ell}]. \quad (13)$$

It should be noticed that this construction slightly differs from the one followed in other studies, where snapshots are centered with respect to the mean value (estimated from the snapshots) and the basis is made

of the dominant eigenvectors of the snapshot covariance matrix [25, 29]. The strategy proposed above was found to provide more accurate results in MD simulations, especially when  $N_a$  becomes large.

For  $0 \leq \ell \leq N_V$ , the Galerkin projection of Eq. (7) expressed in terms of atom displacements (with physical variable  $\mathbf{z}^\ell$ ) reads as

$$[\mathcal{M}_\ell]\ddot{\mathbf{y}}^\ell(t) = \mathcal{F}^\ell(t) \quad (14)$$

where

$$[\mathcal{M}_\ell] = [W_\ell]^\top [M] [W_\ell] \quad (15)$$

is the projected mass matrix and  $\mathcal{F}^\ell$  is the reduced force vector

$$\mathcal{F}^\ell(t) = [W_\ell]^\top \tilde{\mathbf{f}}(t; V_\ell). \quad (16)$$

In general, and as opposed to other frameworks involving nonlinear continuum mechanics formulations, solving for the reduced variable does not allow for a reduction of the computation time. This is due to the fact that most CPU time is spent on assembling  $\tilde{\mathbf{f}}(t; V_\ell)$  and computing  $\mathcal{F}^\ell(t)$ , which is still necessary to evaluate the right-hand side of Eq. (16). It should be noticed that surrogate models can be used to accelerate this intensive task. A discussion about this particular aspect is beyond the scope of this paper; see, *e.g.*, [25] and the references therein.

### 3.2. Stochastic reduced-order model for molecular dynamics

In order to account for model-type uncertainties in the MD formulation, for  $0 \leq \ell \leq N_V$ , the deterministic reduced-order basis  $[W_\ell]$  is replaced by a random matrix  $[\mathbf{W}_\ell]$ , defined on a probability space  $(\Theta, \mathcal{T}, \mathcal{P})$  and with values in the subset

$$\mathcal{S}_{N, n_\ell} = \{[Y] \in \mathbb{M}_{N, n_\ell} \mid [Y]^\top [Y] = [I_{n_\ell}], [B]^\top [Y] = [0_{N_{\text{CD}}, n_\ell}]\} \quad (17)$$

of the compact Stiefel manifold

$$\mathbb{S}_{N, n_\ell} = \{[Y] \in \mathbb{M}_{N, n_\ell} \mid [Y]^\top [Y] = [I_{n_\ell}]\}, \quad (18)$$

where  $\mathbb{M}_{N, n_\ell}$  is the set of all  $N \times n_\ell$  real matrices and  $\mathcal{S}_{N, n_\ell} \subset \mathbb{S}_{N, n_\ell} \subset \mathbb{M}_{N, n_\ell}$ . It follows that  $[\mathbf{W}_\ell]$  satisfies

$$[\mathbf{W}_\ell]^\top [\mathbf{W}_\ell] = [I_{n_\ell}] \quad (19)$$

and

$$[B]^\top [\mathbf{W}_\ell] = [0_{N_{\text{CD}}, n_\ell}] \quad (20)$$

almost surely and can thus be used as a stochastic counterpart for the ROM introduced in Section 3.1. In this work, the probabilistic model for the stochastic reduced-order basis  $[\mathbf{W}_\ell]$  constructed in [23] is deployed. In order to make the paper self-contained, some elements of this construction are provided in Section 3.2.1. The identification of the hyperparameters defining this random matrix is discussed in Section 3.2.3.

#### 3.2.1. Construction of a stochastic reduced-order basis.

In this section, the value of  $\ell$  is fixed in  $\{0, \dots, N_V\}$ . The probabilistic model for the SROB  $[\mathbf{W}_\ell]$  must be constructed such that (i) Eq. (19) and (20) hold almost surely and (ii) the number and dimensionality of model parameters (which are referred to as hyperparameters) remain small enough to make model identification tractable.

In essence, the modeling approach proposed in [23] consists of (i) making the random matrix  $[\mathbf{W}_\ell]$  fluctuate around the deterministic ROB  $[W_\ell]$ , by using the tangent vector space  $T_{[W_\ell]}\mathbb{S}_{N, n_\ell}$  of  $\mathbb{S}_{N, n_\ell}$  at  $[W_\ell] \in \mathbb{S}_{N, n_\ell}$ , and (ii) introducing an appropriate, smooth mapping between this tangent space and the compact Stiefel manifold to ensure that the SROB fulfills Eqs. (19–20) almost surely. Such a construction can be achieved through a sequence of algebraic decompositions, inferred from proper parameterizations of the involved matrix manifolds. These decomposition are detailed hereafter.

Let  $[\mathbf{U}_\ell] = [\mathbf{U}^{\ell,1} \dots \mathbf{U}^{\ell,n_\ell}]$  be a random matrix defined on the probability space  $(\Theta, \mathcal{T}, \mathcal{P})$ , with values in  $\mathbb{M}_{N,n_\ell}$ , and introduce the random matrix

$$[\mathbf{Z}_\ell] = [\mathbf{A}_\ell] - [W_\ell][\mathbf{D}_\ell], \quad (21)$$

where the random matrices  $[\mathbf{A}_\ell]$  and  $[\mathbf{D}_\ell]$  are given by

$$[\mathbf{A}_\ell] = [\mathbf{U}_\ell] - [B][B]^\top[\mathbf{U}_\ell] \quad (22)$$

and

$$[\mathbf{D}_\ell] = \frac{1}{2} ([W_\ell]^\top[\mathbf{A}_\ell] + [\mathbf{A}_\ell]^\top[W_\ell]). \quad (23)$$

It can then be shown that  $[\mathbf{Z}_\ell]$  takes its values in  $T_{[W_\ell]}\mathbb{S}_{N,n_\ell}$  and that  $[\mathbf{A}_\ell]$  satisfies Eq. (20) almost surely. A smooth mapping between  $T_{[W_\ell]}\mathbb{S}_{N,n_\ell}$  and  $\mathbb{S}_{N,n_\ell}$  is then introduced to define  $[\mathbf{W}_\ell]$  as

$$[\mathbf{W}_\ell] = ([W_\ell] + s_\ell[\mathbf{Z}_\ell]) ([I_{n_\ell}] + s_\ell^2[\mathbf{Z}_\ell]^\top[\mathbf{Z}_\ell])^{-1/2}, \quad (24)$$

where  $0 < s_\ell$  is a model parameter controlling the statistical fluctuations of  $[\mathbf{W}_\ell]$ . These fluctuations occur around  $[W_\ell]$  when the random matrix  $[\mathbf{Z}_\ell]$  is centered, which in turn implies that  $[\mathbf{A}_\ell]$  (and thus,  $[\mathbf{U}_\ell]$ ) must be centered as well:

$$E\{[\mathbf{U}_\ell]\} = [0_{N,n_\ell}]. \quad (25)$$

Notice, however, that the mean  $[W_\ell]$  of  $[\mathbf{W}_\ell]$  does not coincide with  $[W_\ell]$ . To provide some modeling flexibility while keeping the parameterization reasonably low-dimensional, the random matrix  $[\mathbf{U}_\ell]$  is written as

$$[\mathbf{U}_\ell] = [\mathbf{G}_\ell][\sigma_\ell], \quad (26)$$

where  $[\sigma_\ell]$  is any matrix in  $\mathbb{M}_{n_\ell}^u$  and where  $[\mathbf{G}_\ell]$  is a random matrix defined on the probability space  $(\Theta, \mathcal{T}, \mathcal{P})$  (see below), with values in  $\mathbb{M}_{N,n_\ell}$ , that satisfies

$$E\{[\mathbf{G}_\ell]\} = [0_{N,n_\ell}] \quad (27)$$

and is constructed such that

$$E\{[\mathbf{G}_\ell]_{jk}[\mathbf{G}_\ell]_{j'k'}\} = [C_{N,\ell}]_{jj'}\delta_{kk'}. \quad (28)$$

It can then be deduced that the fourth-order covariance tensor  $E\{[\mathbf{U}_\ell] \otimes [\mathbf{U}_\ell]\}$  of the random matrix  $[\mathbf{U}_\ell]$  reads as

$$E\{[\mathbf{U}_\ell]_{jk}[\mathbf{U}_\ell]_{j'k'}\} = [C_{N,\ell}]_{jj'}[c_{n,\ell}]_{kk'}, \quad (29)$$

where  $[C_{N,\ell}] \in \mathbb{M}_N^+$  and  $[c_{n,\ell}] = [\sigma_\ell]^\top[\sigma_\ell] \in \mathbb{M}_{n_\ell}^{+0}$  are covariance-like matrices (note that this construction includes the particular case where  $[\sigma_\ell]$  is chosen as the Cholesky factorization of  $[c_{n,\ell}]$ ). It is seen from Eq. (29) that the matrix  $[C_{N,\ell}]$  controls the correlation along the components of the random vector  $\mathbf{U}^{\ell,k}$ ,  $1 \leq k \leq n_\ell$ , while  $[c_{n,\ell}]$  governs the correlation between any pair of random vectors  $\mathbf{U}^{\ell,k}$  and  $\mathbf{U}^{\ell,k'}$ ,  $1 \leq k, k' \leq n_\ell$ .

From the previous derivations, it is seen that the stochastic model can, indeed, be constructed on the random matrix  $[\mathbf{G}_\ell]$ : provided that Eqs. (27) and (28) hold, the use of Eqs. (21), (22), (23) and (26) in Eq. (24) makes the SROB  $[\mathbf{W}_\ell]$  a mathematically consistent randomization of the deterministic ROB  $[W_\ell]$ . It now remains to define the random matrix  $[\mathbf{G}_\ell]$ . As discussed in [23], the definition of the probability law of  $[\mathbf{G}_\ell]$  is somewhat arbitrary in the present context, since the framework is not focused on modeling parametric uncertainties. In this work, the formulation derived in Appendix D of [23] is used and  $[\mathbf{G}_\ell]$  is defined and generated by discretizing independent copies of a second-order, stationary and centered Gaussian random field (see the above reference for details). The covariance function of the latter is parameterized by a scalar parameter  $\beta_\ell > 0$  controlling the rate of decay of the correlation along the rows of  $[\mathbf{G}_\ell]$  (note that the columns of  $[\mathbf{G}_\ell]$  are uncorrelated by construction; see Eq. (28)), and  $[C_{N,\ell}]$  is thus an implicit function of  $\beta_\ell$ ,  $[C_{N,\ell}] = [C_{N,\ell}(\beta_\ell)]$ .



The stochastic model for the SROB is hence parameterized by a maximum of  $2 + n_\ell(n_\ell + 1)/2$  hyperparameters (namely  $s_\ell > 0$ ,  $\beta_\ell > 0$  and the nonzero entries (with a maximum of  $n_\ell(n_\ell + 1)/2$  terms) of the upper triangular matrix  $[\sigma_\ell]$ ), and a minimum of  $2 + n_\ell$  hyperparameters. A strategy to identify these parameters is sketched in the next section. The approach will be exemplified and purposely adapted in Section 4.

### 3.2.2. Definition of the SROM $[\mathbf{W}_\ell]$

For  $0 \leq \ell \leq N_V$ , the SROM is written as

$$\mathbf{Z}^\ell(t) = [\mathbf{W}_\ell] \mathbf{Y}^\ell(t) \quad (30)$$

and

$$[\mathcal{M}_\ell] \ddot{\mathbf{Y}}^\ell(t) = \mathcal{F}^\ell(t), \quad (31)$$

where

$$[\mathcal{M}_\ell] = [\mathbf{W}_\ell]^\top [M] [\mathbf{W}_\ell] \quad (32)$$

and

$$\mathcal{F}^\ell(t) = [\mathbf{W}_\ell]^\top \mathbf{f}(t; V_\ell). \quad (33)$$

In the equation above,  $V_\ell$  is the selected interatomic potential when  $1 \leq \ell \leq N_V$ . For  $\ell = 0$ , the potential  $V_0$  represents a reference potential in the family of potentials under consideration, denoted by  $V_{\ell^*}$ , with  $1 \leq \ell^* \leq N_V$ .

### 3.2.3. Identification of the SROM $[\mathbf{W}_\ell]$

Throughout this section, the value of  $\ell$  is fixed in  $\{0, \dots, N_V\}$  and the dependence of the hyperparameters on  $\ell$  is omitted to simplify the presentation. In general, the definition of a strategy to identify the parameters of the SROB is problem specific and typically based on the selection of a quantity of interest (QoI). Regardless of the latter, it is reasonable to assume that the (potentially time-dependent) mean and variance of the QoI constitute target properties. This allows us to derive the following prototypical form for the cost function:

$$J(\boldsymbol{\alpha}) = w J_m(\boldsymbol{\alpha}) + (1 - w) J_v(\boldsymbol{\alpha}), \quad (34)$$

where  $\boldsymbol{\alpha}$  symbolically gathers all hyperparameters ( $\boldsymbol{\alpha} = (s, \beta, [\sigma])$ ),  $J_m$  and  $J_v$  are normalized relative errors in mean and variability (which can be expressed in terms of variance or confidence interval) between some reference dataset and results predicted with the SROM (see Section 4 for examples), and  $w \in [0, 1]$  is a weight allowing the two contributions to be balanced. The optimal solution is then defined as

$$\boldsymbol{\alpha}^{\text{opt}} = \arg \min_{\boldsymbol{\alpha} \in \mathcal{C}_\alpha} J(\boldsymbol{\alpha}), \quad (35)$$

where the admissible set is taken, in this paper, as  $\mathcal{C}_\alpha = \mathbb{R}_{>0} \times \mathbb{R}_{>0} \times \mathbb{M}_n^{+u}$  (see Section 3.2.1). The choice of the reference dataset is goal-oriented. For instance, if this reference is taken as the prediction of the QoI obtained through the HFM (which is the non-reduced molecular dynamics simulation in this study), then the SROM aims at accounting for the modeling error generated by the reduction (see Section 4.2 for an application). If some variability is introduced in the HFM using different interatomic potentials, the SROM is intended to incorporate both the reduction error and model-type uncertainties. This case will be specifically addressed in Section 4.3. Finally, the reference dataset can be composed of experimental data, in which case the SROM integrates model reduction error, model-form uncertainties and possible parametric uncertainties.

The optimization problem given by Eq. (35) is nonconvex and generally high-dimensional when  $[\sigma] \in \mathbb{M}_n^{+u}$ . It should be noticed that the dimension of this problem can be reduced if  $[\sigma]$  is sought in  $\mathbb{M}_n^u$  (see [30]). In order to solve this problem at a reasonable computational cost, a multi-step methodology is pursued as follows [23]. In the description below, variables that are not active at a given step are assigned the values determined from the previous step.

Step 1. Active variables:  $(s, \beta, \sigma)$ , where  $\sigma$  is such that  $[\sigma] = \sigma[I_n]$ .

An optimal point that minimizes the cost function  $J$  over a tensor product grid on the parameter space is sought. The following support restrictions are applied:  $0 < s_\ell \leq s \leq s_u$ ,  $0 < \beta_\ell \leq \beta \leq \beta_u$  and  $0 < \sigma_\ell \leq \sigma \leq \sigma_u$ . The solution is denoted by  $\alpha_1 = (s_1, \beta_1, [\sigma_1])$ , with  $[\sigma_1] = \sigma_1[I_n]$ .

Step 2. Active variables:  $(s, \beta, [\sigma])$ , where  $[\sigma]$  is a diagonal matrix.

The interior point method is used to solve Eq. (35), considering the inequality constraints from Step 1 and starting from  $\alpha_1$ . The solution is denoted by  $\alpha_2 = (s_2, \beta_2, [\sigma_2])$ .

Step 3. Active variable:  $[\sigma^*]$ , where  $[\sigma^*]$  is an upper triangular matrix with null diagonal such that  $[\sigma] = [\sigma_2] + [\sigma^*]$ .

The interior point method is used to solve Eq. (35), where no constraint is imposed on the entries of  $[\sigma^*]$ . The solution is denoted by  $\alpha_3 = (s_2, \beta_2, [\sigma_3])$ , with  $[\sigma_3] = [\sigma_2] + [\sigma_3^*]$ .

Step 4. Active variable:  $\beta$ .

The optimal value  $\beta^{\text{opt}}$  of  $\beta$  is found by minimizing the cost function  $J$  over a uniform discretization of  $[\beta_\ell, \beta_u]$ . The solution is denoted by  $\alpha_4 = (s_2, \beta^{\text{opt}}, [\sigma_3])$ .

Step 5. Active variables:  $(s, [\sigma_d])$ , where  $[\sigma_d]$  is a diagonal matrix such that  $[\sigma] = [\sigma_d] + [\sigma_3^*]$ .

The interior point method is used to solve Eq. (35) and find the optimal values  $s^{\text{opt}}$  and  $[\sigma_d^{\text{opt}}]$ . The optimal solution is then defined as  $\alpha^{\text{opt}} = (s^{\text{opt}}, \beta^{\text{opt}}, [\sigma^{\text{opt}}])$ , with  $[\sigma^{\text{opt}}] = [\sigma_d^{\text{opt}}] + [\sigma_3^*]$ .

Note that the lower and upper bounds in the above steps are problem specific.

## 4. Applications

### 4.1. Overview

In this section, we assess the relevance of the stochastic reduced-order model to represent, and eventually to propagate, model-form uncertainties in MD simulations. A graphene sheet system is selected as a prototypical application, and the following set of potentials (available in LAMMPS) will be considered:

- $V_1$ : modified Morse [25];
- $V_2$ : Airebo [3];
- $V_3$ : BOP [31];
- $V_4$ : LCBOP [4];
- $V_5$ : REBOOII [32];
- $V_6$ : Tersoff-2010 [12].

The values of the parameters used for each potential can be found in the references above.

The case of uncertainties raised by *model reduction* is first addressed in Section 4.2 for  $N_V = 1$ , with  $V_1$  corresponding to a modified Morse potential (application 1). The objective is to identify the associated SROB  $[\mathbf{W}_1]$  modeling the uncertainties generated by the use of the ROM instead of the HFM.

Model uncertainties generated by the *selection of potentials* is then discussed in Section 4.3, including the effect of the truncation error (application 2). In this case, we consider the six different interatomic potentials listed above,  $N_V = 6$ . A SROB  $[\mathbf{W}_1]$  is constructed by randomizing the ROB  $[W_1]$ , associated with a modified Morse potential  $V_1$ , and the SROM is identified to account for the variability induced by the use of the family of potentials  $\{V_1, \dots, V_6\}$ .

In the last application, discussed in Section 4.4, we investigate the use of a SROB  $[\mathbf{W}_4]$ , associated with the LCBOP potential  $V_4$  (application 3-a), and a SROB  $[\mathbf{W}_0]$  constructed by considering snapshots associated with all potentials  $\{V_1, \dots, V_6\}$  (application 3-b). In both cases, the SROM is identified to

represent the *model error between molecular dynamics computations and a continuum-mechanics model* calibrated with experimental data.

All MD simulations reported in this section are performed in LAMMPS with a time step of 1 fs. As previously indicated, the SROMs were also implemented in LAMMPS.

#### 4.2. Application 1: quantification of model uncertainties induced by order reduction

In this first application, the 272-atom graphene sheet system introduced in Section 2.2 is considered (with  $A = 6 \text{ kcal.mol}^{-1}.\text{\AA}^{-1}$  and  $\omega = 20 \text{ GHz}$ ), given the high sensitivity of the results to interatomic potentials (see the parametric analysis in the aforementioned section). A modified Morse potential [1] is used for the construction of the deterministic reduced-order basis.

##### 4.2.1. Construction of the deterministic ROB

In order to construct the deterministic reduced-order basis  $[W_1]$ , the graphene sheet is subjected to a 300K thermo-equilibrium and snapshots are collected every 1,000 fs. The number of snapshots necessary to construct the ROB is determined by analyzing the convergence of the eigenvalues of the empirical estimate of the second-order tensor moment of  $\boldsymbol{\mu}$ , using the function  $N_s(1) \mapsto \text{Tr}([S(N_s(1))]^2)/N_s(1)$  (see Eq. (10)) and the value  $N_s(1) = 1,000$  is selected (with snapshots uniformly sampled over a period of the excitation). Three snapshots are shown in Fig. 4.

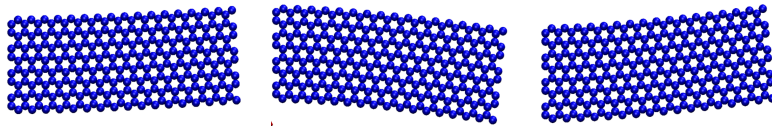


Figure 4: Examples of snapshots used to construct the reduced-order basis  $[W_1]$ .

The dimension  $n_1$  of the ROB can be determined in a standard manner, by characterizing the convergence of the function

$$\varepsilon(n_1) = 1 - \frac{\sum_{i=1}^{n_1} [S(N_s(1))]_{ii}^2}{\text{Tr}([S(N_s(1))]^2)}, \quad (36)$$

in which  $N_s(1)$  was determined from the convergence analysis ( $N_s(1) = 1,000$ ), and selecting  $n_1 < N_s(1)$  such that  $\varepsilon(n_1) \leq \varepsilon$ . The graph of the function  $n_1 \mapsto \varepsilon(n_1)$  is shown in Fig. 5. In this application, the value  $\varepsilon = 10^{-4}$  is chosen, which corresponds to  $n_1 = 10$ , and computations are initialized with a non-stress-free configuration. A comparison between the HFM (standard molecular dynamics) and ROB-based predictions for the  $\mathbf{e}^2$ -displacement is displayed in Fig. 6. The discrepancies observed between the HFM and ROB-based results indicate that the ROM cannot fully capture the nonlinear response of the system, despite the use of a small value for the error parameter  $\varepsilon$ . In the next section, the SROM is used to account for this reduction-induced error.

##### 4.2.2. Stochastic reduced-order molecular dynamics

In order to deploy the SROM, the hyperparameters defining the stochastic reduced-order basis  $[\mathbf{W}_1]$  must be identified on a reference dataset. As discussed earlier, the definition of this dataset and of the cost function  $J$  is goal-oriented. Following the results in Section 4.2.1 (see Fig. 6), the  $\mathbf{e}^2$ -displacement at time  $t^* = 60,000$  fs is selected as the QoI, and the cost function is constructed to impose a match between the HFM result and the mean of the SROM-based predictions. Fluctuations in the stochastic model are calibrated to account for the fluctuations of the HFM outcomes around the mean response of the SROM.

Let  $\mathbf{y}^{\text{HFM}}$  and  $\mathbf{y}^{\text{ROM}}$  be the vectors in  $\mathbb{R}^{N_a}$  (with  $N_a = 272$ ) gathering the  $\mathbf{e}^2$ -displacements of all atoms at time  $t^*$ , obtained with the HFM and the ROM, respectively. Denote by  $\mathbf{Y}^{\text{SROM}}(\boldsymbol{\alpha}; V_1)$  the random vector defined in  $(\Theta, \mathcal{T}, \mathcal{P})$ , with values in  $\mathbb{R}^{N_a}$ , corresponding to the predictions (expressed in physical space) at

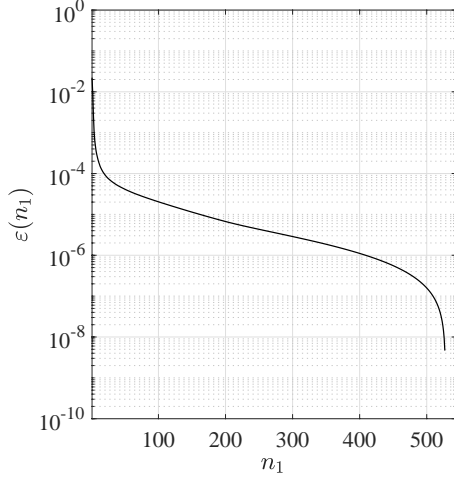


Figure 5: Plot of the error function  $n_1 \mapsto \varepsilon(n_1)$ .

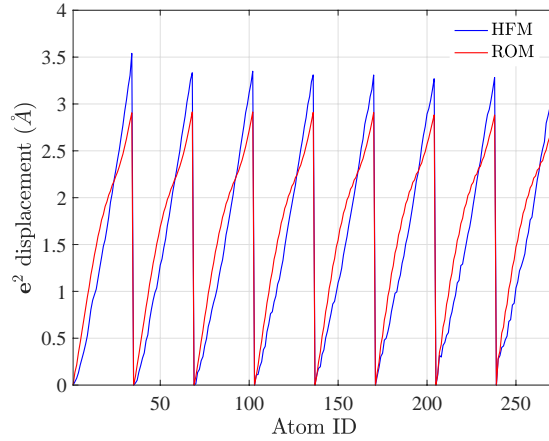


Figure 6: HFM and ROM-based predictions for the  $\mathbf{e}^2$ -displacement at  $t^* = 60,000$  fs.

time  $t^*$  with the SROB  $[\mathbf{W}_1]$ . The mean-related contribution in the cost function  $J$  (see Eq. (34)) is then defined as

$$J_m(\boldsymbol{\alpha}) = \|\mathbf{y}^{\text{HFM}} - E\{\mathbf{Y}^{\text{SROM}}(\boldsymbol{\alpha}; V_1)\}\|^2 / \|\mathbf{y}^{\text{HFM}}\|^2, \quad (37)$$

while the contribution related to fluctuations is taken as

$$J_v(\boldsymbol{\alpha}) = \|\mathbf{v}^{\text{HFM-ROM}} - \mathbf{v}^{\text{SROM}}(\boldsymbol{\alpha}; V_1)\|^2 / \|\mathbf{v}^{\text{HFM-ROM}}\|^2, \quad (38)$$

where  $\mathbf{v}^{\text{HFM-ROM}}$  is a  $\gamma$ -weighted relative error between the HFM and ROM results,  $v_i^{\text{HFM-ROM}} = \gamma|y_i^{\text{HFM}} - y_i^{\text{ROM}}|$  for  $1 \leq i \leq N_a$  (with  $\gamma \geq 0$ ), and the components of  $\mathbf{v}^{\text{SROM}}(\boldsymbol{\alpha}; V_1)$  are  $v_i^{\text{SROM}}(\boldsymbol{\alpha}; V_1) = \text{Var}\{Y_i^{\text{SROM}}(\boldsymbol{\alpha}; V_1)\}$  for  $1 \leq i \leq N_a$ . The hyperparameters defining the SROM are identified following the steps presented in Section 3.2.3, with  $w = 0.99$  (notice that this value allows  $J_m$  and  $J_v$  to have different contributions in the global cost function). The  $\mathbf{e}^2$ -displacement for all atoms at time  $t^*$  is shown in Fig. 7 for the HFM, the (deterministic) ROM and the SROM. The confidence interval is estimated with 1,000 independent realizations of the SROM.

It is seen that the mean value of the SROM computation closely replicates the result obtained with the HFM, regardless of the value of  $\gamma$ . The effect of the latter on the spread of the confidence interval can also

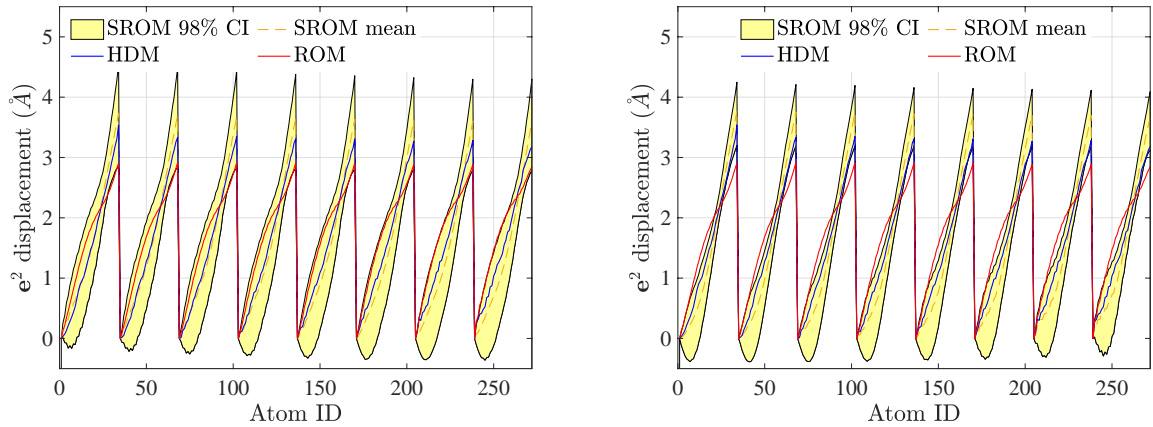


Figure 7: Vertical displacement for all atoms at time  $t^*$ : HFM (blue line), ROM-based (red line) and SROM-based (orange line; yellow region: 98% confidence interval) results. The left and right panels correspond to  $\gamma = 1$  and  $\gamma = 0.5$ , respectively.

be observed. In particular, the variability exhibited by the QoI decreases with  $\gamma$ , with a plateau observed for some minimal value of the parameter. In the present case,  $\gamma = 0.5$  yields a reasonable approximation in the sense that (i) the mean obtained with the SROM is close of the HFM result, and (ii) the HFM prediction is included in the confidence region associated with the identified SROM. The effect on the total energy is next illustrated in Fig. 8. In this figure, ROM-based results are presented for several values of  $n$ . Note that the construction of SROMs for all these values of  $n$  was not pursued, since the number of hyperparameters raises up significantly.

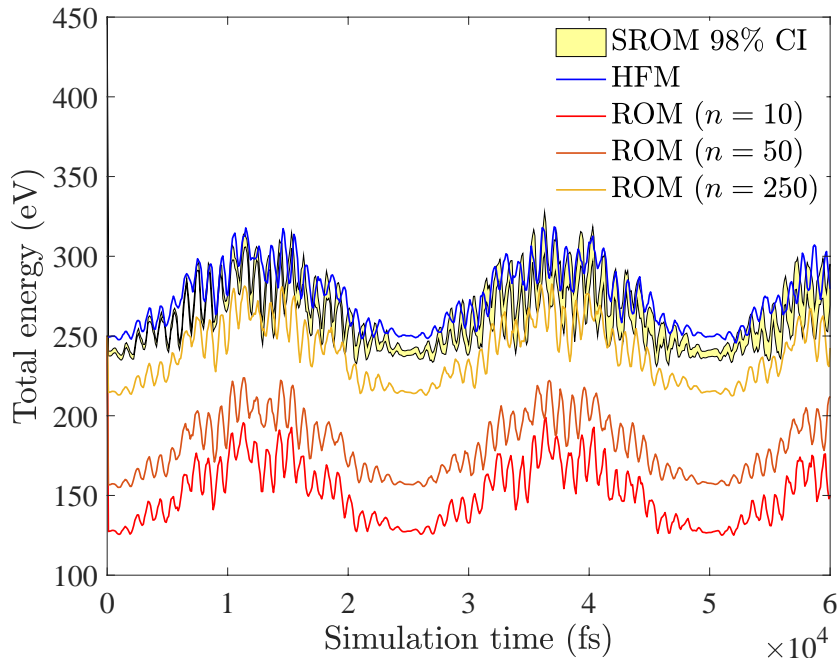


Figure 8: Evolution of the total energy for the HFM, the ROM with  $n \in \{10, 50, 250\}$ , and the identified SROM.

In addition, the Gaussian kernel density estimation and confidence region for the interatomic bond length,

obtained with the HFM, ROM, and SROM, are shown in Fig. 9. The discrepancies between the HFM and

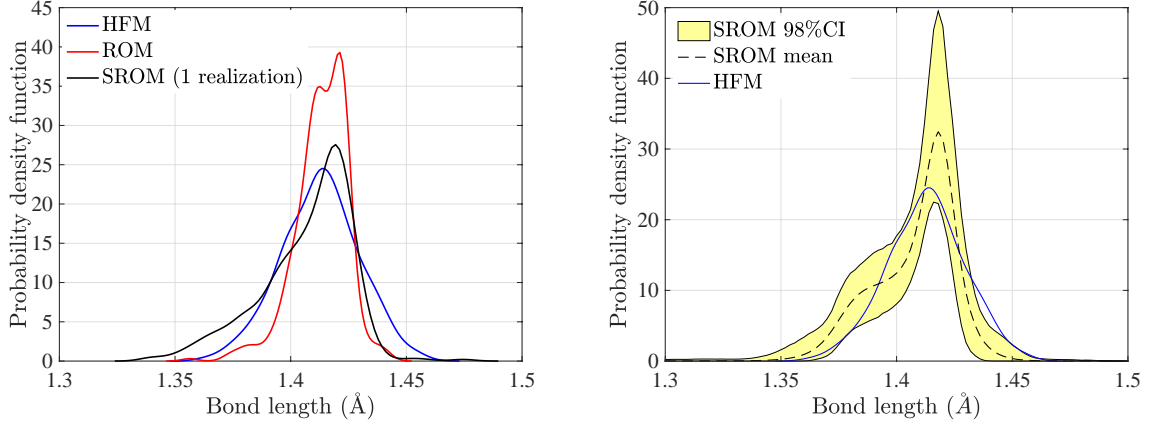


Figure 9: Gaussian kernel density estimator of the probability density function for the bond length for the HFM, ROM and SROM (left), and confidence interval for the probability density function estimated with the SROM  $[\mathbf{W}_1]$  (right).

the ROM, observed for both the energy and the probability density function of the bond length, are due to the value of  $n$ , which is not sufficiently large. It should be noticed that  $n$  could have been chosen greater (at the expense of additional computational cost): the selected value is, however, sufficient to construct the SROM, which still delivers a good approximation for the energy (see Fig. 8). Moreover, it is seen that the probability density function estimated for one realization of the SROM is close to the probability density function obtained with the HFM, and that the latter is, indeed, included in the 98% confidence region of the SROM. This demonstrates the ability of the SROM to generate a coherent confidence region, even when the ROM has not converged to the HFM.

#### 4.3. Application 2: quantification of model uncertainties due to interatomic potential selection

We now turn to the assessment of model uncertainties induced by interatomic potentials. To that end, the graphene sheet model is still considered and six widely used potentials (namely, the Airebo, BOP, LCBOP, modified-Morse, REBOII, and Tersoff-2010 potentials) are considered to generate a spread in the HFM results. The predicted  $\mathbf{e}^2$ -displacement at a given time instance is shown in Fig. 10. It is seen that the choice of the potential has a significant impact on the predictions, with a maximum difference over all atoms equal to 59.7% (note that this relative error is normalized by the average between the maximum and minimum values). More specifically, it is observed that the BOP potential predicts the most compliant response, while the Tersoff-2010 potential yields the most rigid response (under the same boundary conditions).

In the present application, the modified Morse potential is used to construct the ROB  $[\mathbf{W}_1]$ . In order to identify the parameters in the SROB  $[\mathbf{W}_1]$ , the definition of the cost function  $J$  is adapted such that the minimum and maximum envelopes of the response taken over all potentials are contained within the 98% confidence region of the SROM. This constitutes an alternative to the choice given by Eqs. (37) and (38). We adopt the same notations as in Section 4.2.2 for the quantity of interest (that is, the  $\mathbf{e}^2$ -displacement of all atoms at  $t^*$ ), and let  $\mathbf{y}_-^{\text{HFM}}$  and  $\mathbf{y}_+^{\text{HFM}}$  be the vectors of lower and upper envelopes over the potentials,  $(y_-^{\text{HFM}})_i = \min(\{y_i(V_j)\}_{j=1}^{N_V})$  and  $(y_+^{\text{HFM}})_i = \max(\{y_i(V_j)\}_{j=1}^{N_V})$ , where  $N_V$  is the number of potentials considered in the HFM ( $N_V = 6$  in the present application). In addition, we introduce the two vectors  $\mathbf{q}_-^{\text{SROM}}(V_1)$  and  $\mathbf{q}_+^{\text{SROM}}(V_1)$  in  $\mathbb{R}^{N_a}$ , the components of which correspond to the lower and upper values of the 98% confidence region of the components of  $\mathbf{Y}^{\text{SROM}}(\boldsymbol{\alpha}; V_1)$ :  $(q_-^{\text{SROM}}(V_1))_i = P_{98\%}^-(Y_i^{\text{SROM}}(\boldsymbol{\alpha}; V_1))$  and  $(q_+^{\text{SROM}}(V_1))_i = P_{98\%}^+(Y_i^{\text{SROM}}(\boldsymbol{\alpha}; V_1))$ . We then take  $J(\boldsymbol{\alpha}) = J_v(\boldsymbol{\alpha})$ , with

$$J_v(\boldsymbol{\alpha}) = (\|\mathbf{y}_+^{\text{HFM}} - \mathbf{q}_+^{\text{SROM}}(V_1)\|^2 + \|\mathbf{y}_-^{\text{HFM}} - \mathbf{q}_-^{\text{SROM}}(V_1)\|^2) / \|\mathbf{y}_+^{\text{HFM}} - \mathbf{y}_-^{\text{HFM}}\|^2. \quad (39)$$

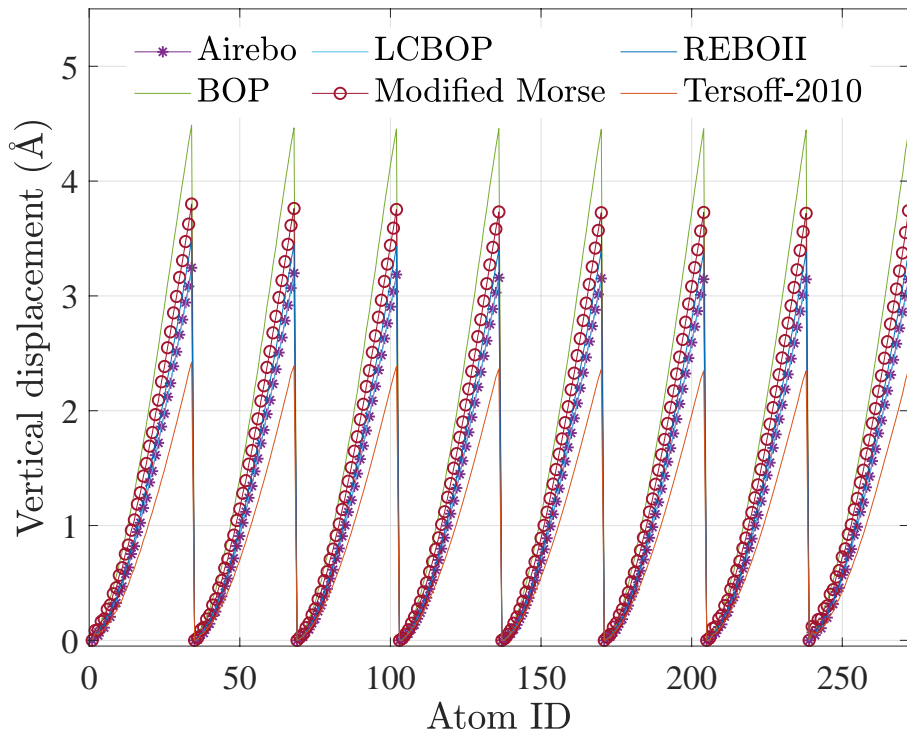


Figure 10: Vertical displacement at  $t^* = 60,000$  fs, predicted by the HFM with different interatomic potentials.

The SRM is calibrated pursuing the multi-step approach (see Section 3.2.3), and the comparison between the HFM outcomes (for several potentials) and the SRM is shown in Figs. 11 and 12. It is seen that the SRM allows the variability generated with the considered set of potentials to be accurately reproduced for all atoms of the graphene sheet.

#### 4.4. Application 3: quantification of model uncertainties based on a continuum model

In this final application, HFM predictions obtained for different potentials are gauged against results obtained by a hyperelastic continuum model. The latter was specifically constructed and identified in [33], showing a good agreement with the AFM experiments conducted in [34]. It is used below to generate a reference dataset under uniaxial loading conditions, referred to as the continuum model. The SRM is then deployed to represent the variability induced by the choice of an interatomic potential.

##### 4.4.1. Molecular dynamics setup

A  $41.74 \text{ \AA} \times 42.54 \text{ \AA}$  graphene sheet is built with periodic boundary conditions as shown in Fig. 13. The graphene model includes 680 carbon atoms, which is large enough to make the response of the system size independent. An interatomic bond distance of  $1.418 \text{ \AA}$ , widely accepted in the literature, is adopted [35].

In order to construct an appropriate ROB, two MD simulations are performed: the graphene model is subjected to uniaxial-strain deformation (Dirichlet boundary conditions), first along the zigzag direction, and then along the armchair direction (see the right panel in Fig. 13). Periodic Dirichlet boundary conditions are then applied, and a classical relaxation is imposed after each increment of the boundary displacement. The boundary displacement is imposed through a stepping scheme with increments imposed every 100 fs (leading to  $N_\epsilon$  increments of the engineering strain), and the sheet is stretched up to a strain of 0.3, for  $T = 30,000$  fs. A time step of 1 fs is used and all the MD simulations are performed under the NVE

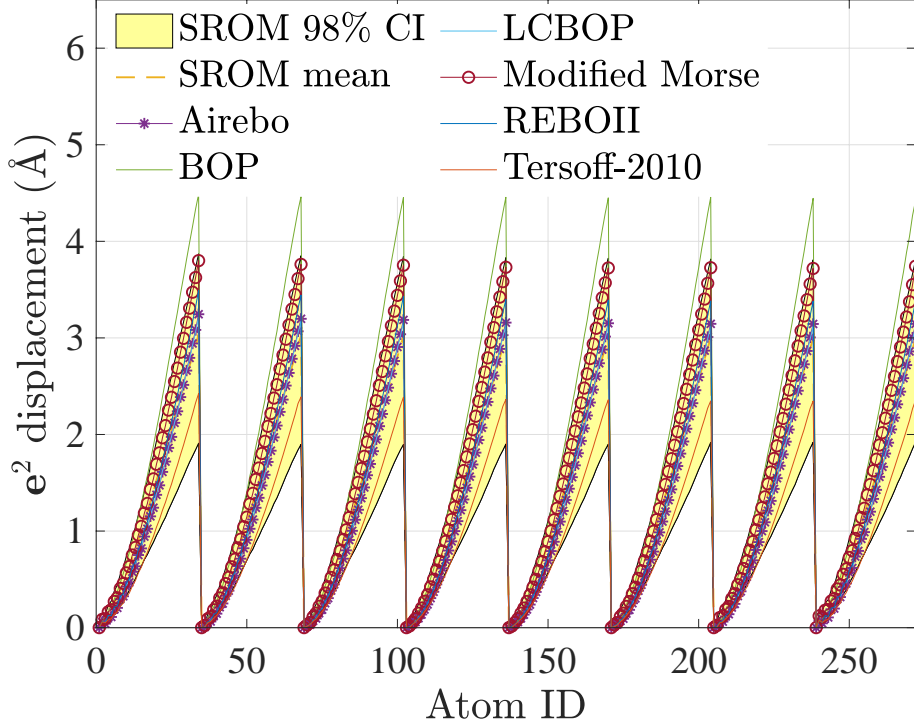


Figure 11: Vertical displacement of all atoms at  $t^* = 60,000$  fs, predicted by the HFMs, and associated 98% confidence interval estimated with the SROM (from 1,000 independent samples).

ensemble. Five interatomic potentials are considered in our study, namely the Airebo [3], BOP [31], LCBOP [4], REBOII [32] and Tersoff-2010 [12].

Following [33], the strain energy under uniaxial tension along zigzag and armchair directions (respectively denoted with underscripts  $zz$  and  $ac$ ) can be expressed as

$$u_{zz}(\epsilon) = \frac{1}{2} \frac{E}{1-v^2} \epsilon^2 + \frac{1}{6} C_{111} \epsilon^3 \quad (40)$$

and

$$u_{ac}(\epsilon) = \frac{1}{2} \frac{E}{1-v^2} \epsilon^2 + \frac{1}{6} C_{222} \epsilon^3, \quad (41)$$

where  $E = 312 \text{ Nm}^{-1}$  is the Young's modulus,  $v = 0.31$  is the Poisson ratio,  $C_{111} = -1689.2 \text{ Nm}^{-1}$  and  $C_{222} = -1487.7 \text{ Nm}^{-1}$  are two elastic constants and  $\epsilon$  is the engineering strain. Fig. 14 shows the evolution of the strain energy as predicted by either the continuum model above (which matches the experimental results) or by molecular dynamics simulations, parameterized by the different potentials. The model-type uncertainties can clearly be visualized, and it is seen that none of the interatomic potentials captures the nonlinear behavior of the graphene system predicted by the continuum model.

#### 4.4.2. Application 3-a: stochastic reduced-order modeling with SROB [W<sub>4</sub>]

The ROB is constructed with the concatenation of 1,000 snapshots in each direction (hence, a total of 2,000 snapshots), obtained with the LCBOP potential  $V_4$ . Following an analysis of convergence, a reduced dimension  $n_4 = 10$  is used. Let  $\mathbf{u}^{\text{CM}}$  be the vector in  $\mathbb{R}^{2N_\epsilon}$  gathering the values of the potential energy (strain energy) at all engineering strain increments along the two (armchair and zigzag) directions, obtained



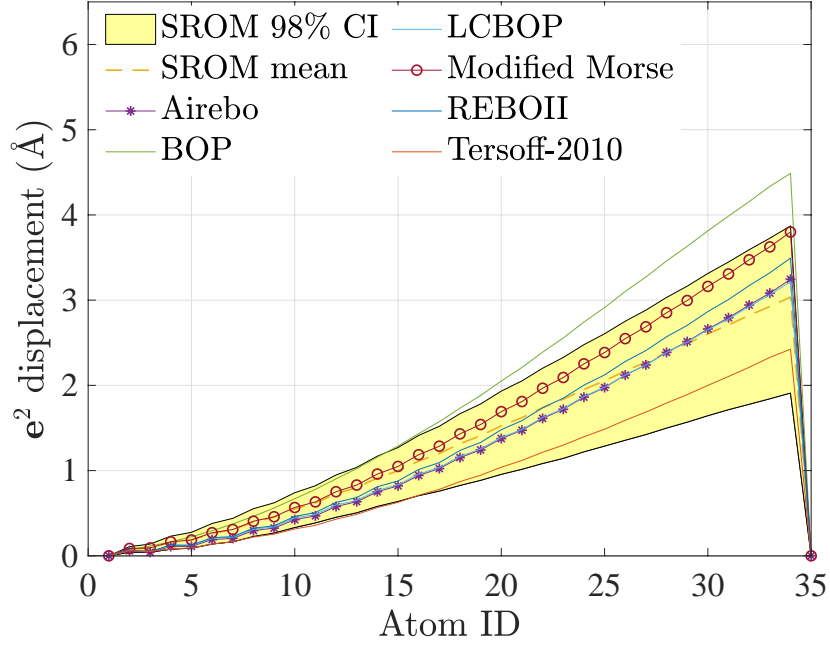


Figure 12: Vertical displacement of a group of atoms at  $t^* = 60,000$  fs, predicted by the HFMs, and associated 98% confidence interval estimated with the SROM (from 1,000 independent samples).

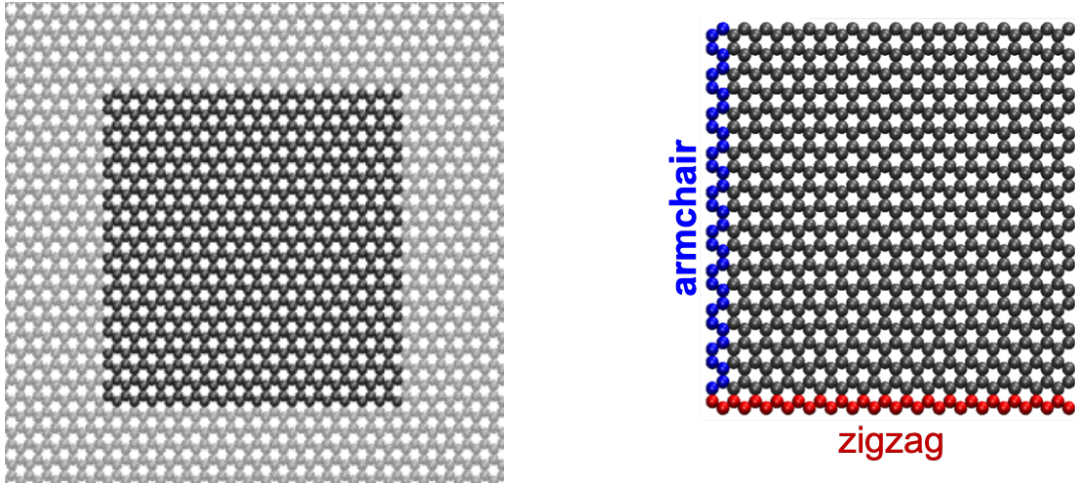


Figure 13: Left panel: graphene sheet model with periodic boundary conditions (gray atoms are periodic images). Right panel: definition of the armchair and zigzag directions.

with the experimentally-consistent continuum model (denoted as CM). The hyperparameters are identified by using the cost function defined in Eq. (34),  $J(\boldsymbol{\alpha}) = wJ_m(\boldsymbol{\alpha}) + (1-w)J_v(\boldsymbol{\alpha})$ , in which

$$J_m(\boldsymbol{\alpha}) = \|\mathbf{u}^{\text{CM}} - E\{\mathbf{U}^{\text{SROM}}(\boldsymbol{\alpha}; V_4)\}\|^2 / \|\mathbf{u}^{\text{CM}}\|^2, \quad (42)$$

and

$$J_v(\boldsymbol{\alpha}) = \|\mathbf{v}^{\text{CM-ROM}} - \mathbf{v}^{\text{SROM}}(\boldsymbol{\alpha}; V_4)\|^2 / \|\mathbf{v}^{\text{CM-ROM}}\|^2, \quad (43)$$

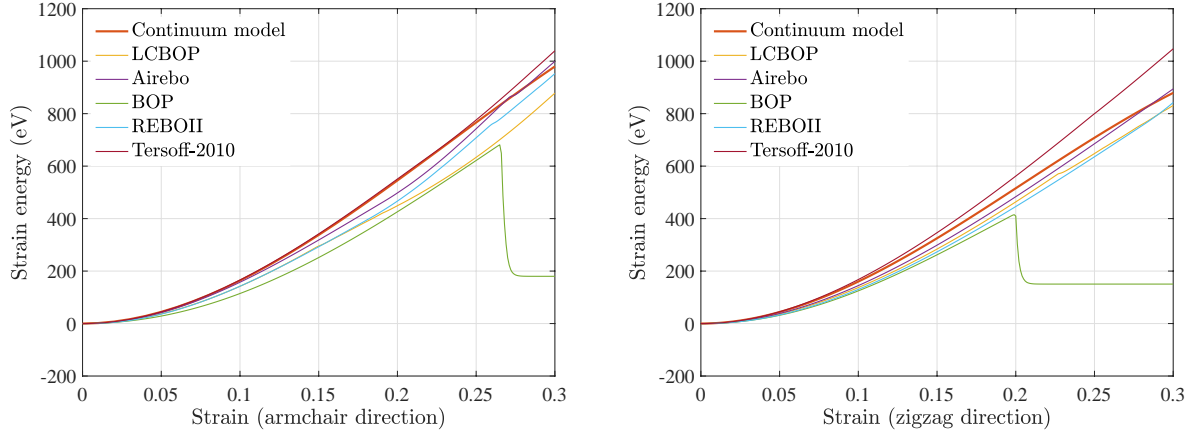


Figure 14: Evolution of the strain energy predicted by the continuum-mechanics model and different molecular dynamics systems. The left and right panels correspond to the armchair and zigzag directions, respectively.

where  $\mathbf{U}^{\text{SROM}}(\boldsymbol{\alpha}; V_4)$  is the concatenated random vector with values in  $\mathbb{R}^{2N_\epsilon}$ , corresponding to the two MD predictions (along the two directions) of the potential energy obtained with the SROB  $[\mathbf{W}_4]$  for all strain increments, the components of  $\mathbf{v}^{\text{CM-ROM}}$  are defined as  $v_i^{\text{CM-ROM}} = \gamma |u_i^{\text{CM}} - u_i^{\text{ROM}}|$  for  $1 \leq i \leq 2N_\epsilon$  (with  $\gamma = 0.5$ ), and the components of  $\mathbf{v}^{\text{SROM}}(\boldsymbol{\alpha}; V_4)$  are given by  $v_i^{\text{SROM}}(\boldsymbol{\alpha}; V_4) = \text{Var}\{U_i^{\text{SROM}}(\boldsymbol{\alpha}; V_4)\}$  for  $1 \leq i \leq 2N_\epsilon$ .

Results obtained with the identified model are shown in Figs. 15 and 16. It is seen in Fig. 15 that the

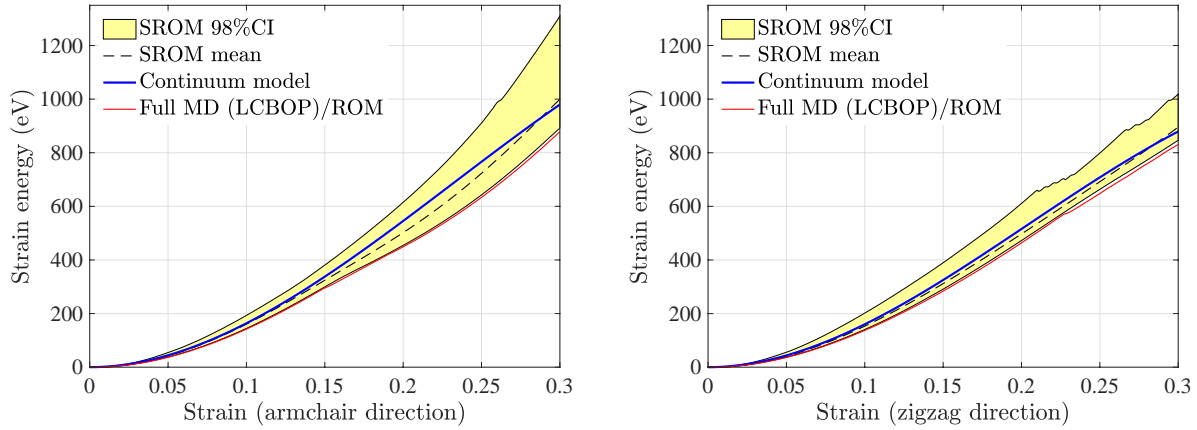


Figure 15: Confidence region and evolution of the potential energy function in the armchair and zigzag directions: continuum-mechanics model and  $[\mathbf{W}_4]$  SROM predictions.

SROM allows the mean model to approach the continuum-mechanics target with a reasonably good accuracy, regardless of the loading direction. For the specific case of the armchair direction, the use of the SROB  $[\mathbf{W}_4]$ , constructed with the snapshots associated with  $V_4$  only, leads to an increase of the variability. This suggests that the use of a SROB  $[\mathbf{W}_0]$ , constructed with the concatenation of snapshots obtained for the family of potentials, would decrease the level of uncertainties, that is, the width of the confidence region: this point will be discussed in detail in Section 4.4.3. In terms of quality assessment, it should be noted at this stage that while the MD results associated with all potentials are not considered in the identification of the SROM, the latter is still able to accurately reproduce the uncertainties generated by the considered

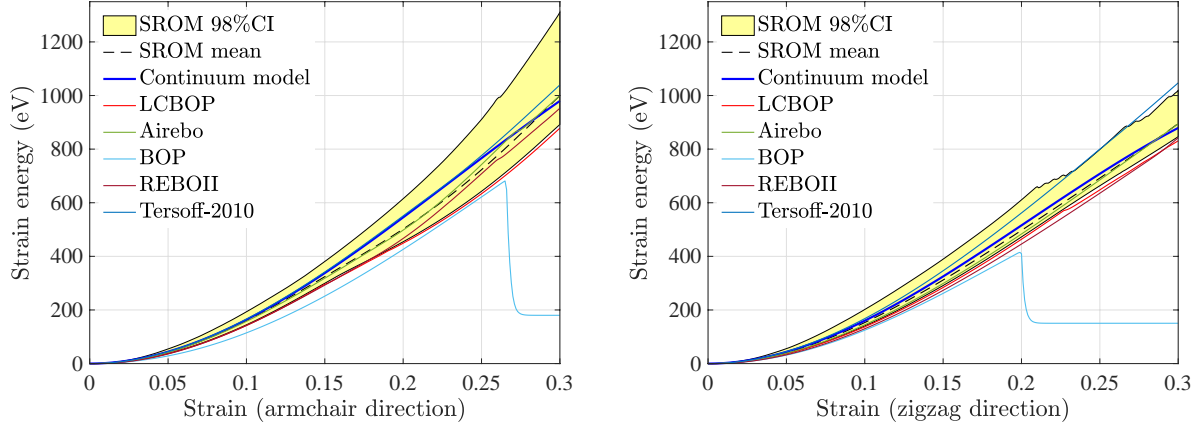


Figure 16: Confidence region and evolution of the potential energy function in the armchair and zigzag directions: continuum-mechanics model, molecular dynamics results for all potentials and SROM predictions.

family of potentials.

The confidence interval on the bond length is finally shown in Fig. 17, in both the armchair and zigzag directions. It is seen that the estimated confidence interval is very sensitive to the direction under which

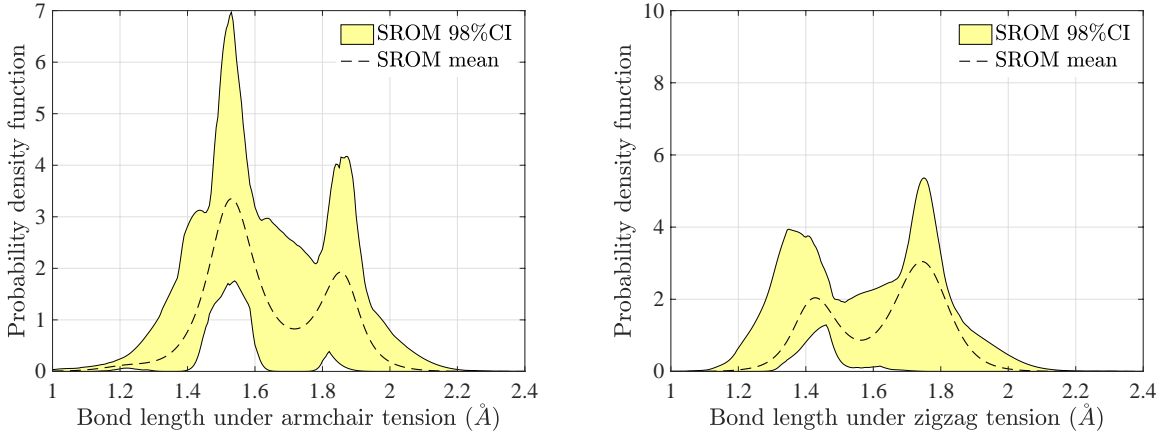


Figure 17: Confidence interval for the probability density function of the bond length, estimated with the SROB  $[\mathbf{W}_4]$  in the armchair (left) and zigzag (right) directions.

loading is applied. This property is due to the very structure of the graphene sheet, depicted in Fig. 13 (note, in particular, that the mechanical response is not invariant with respect to  $\pi/2$  rotations). In addition, the variability exhibited by the probability density functions for the bond length is rather large, which is due to the spread observed with the identified  $[\mathbf{W}_4]$  SROM (see Fig. 15). The impact of the latter can also be noticed on the mean bond length, estimated as a function of the applied engineering strain; see Fig. 18.

#### 4.4.3. Application 3-b: stochastic reduced-order modeling with SROB $[\mathbf{W}_0]$

The SROM defined with the SROB  $[\mathbf{W}_0]$ , considering  $V_4$  as the reference potential ( $\ell^* = 4$ ), is finally deployed. The results obtained in this case are shown in Fig. 19 and must be compared with those presented in Fig. 15. It is seen that the SROM with  $[\mathbf{W}_0]$  allows us to predict the target (which is the experimentally-consistent continuum-mechanics model) with a level of uncertainties that is significantly smaller than the one

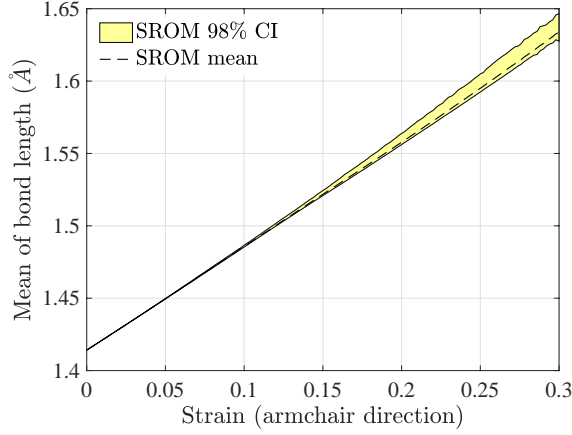


Figure 18: Evolution of the confidence interval for the mean bond length (with SROB  $[\mathbf{W}_4]$ ), as a function of applied engineering strain (armchair direction).

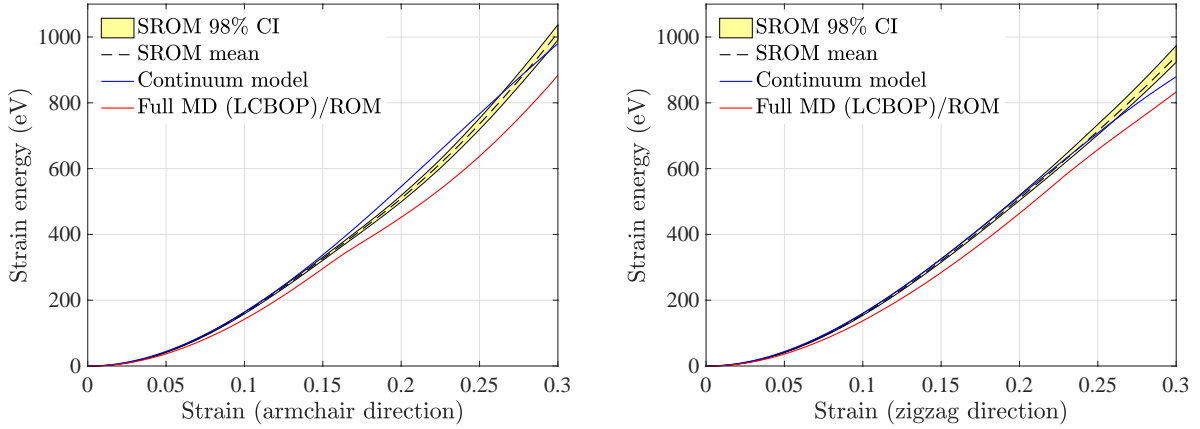


Figure 19: Confidence region and evolution of the potential energy function in the armchair and zigzag directions: continuum-mechanics model and  $[\mathbf{W}_0]$  SROM predictions.

for the SROM with  $[\mathbf{W}_4]$ . More precisely, the level of variation at the largest strain ( $\varepsilon = 0.3$ ) is decreased from 51.33% to 6.51% in the armchair direction, and from 24.17% to 10% in the zigzag direction. Quite remarkably, the SROM can predict the target with a confidence region that does not include the ROM response. This is an important difference with the results presented for the SROM defined with  $[\mathbf{W}_4]$ , in which the ROM results are included in the confidence region for both the armchair and zigzag directions (see Fig. 15) – at the expense of a larger confidence region.

Confidence intervals on the bond length are shown in Fig. 20 for the two loading directions. By comparing this figure with Fig. 17, where the confidence intervals were estimated with the SROB  $[\mathbf{W}_4]$ , it is seen that the use of the SROB  $[\mathbf{W}_0]$  allows the experimentally-consistent continuum model to be approximated with much less variability. This conclusion can also be drawn from Fig. 21, where the mean bound length is represented as a function of the applied strain. The lower and upper bounds of the confidence interval for the mean bound length are almost indistinguishable, which is in contrast with the result obtained with the SROB  $[\mathbf{W}_4]$  (see Fig. 18). These results demonstrate the robustness of the SROM, especially when constructed with the whole family of potentials.

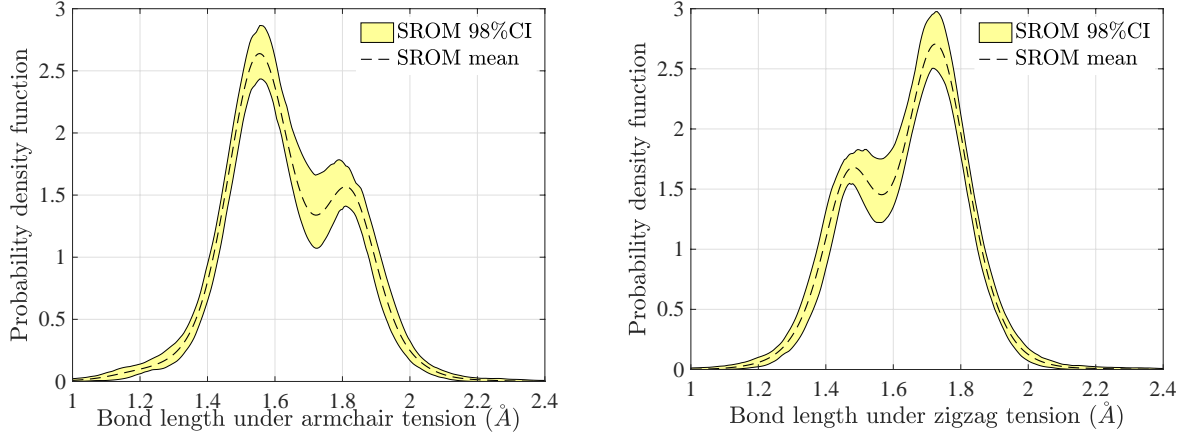


Figure 20: Confidence interval for the probability density function of the bond length, estimated with the SROB  $[\mathbf{W}_0]$  in the armchair (left) and zigzag (right) directions..

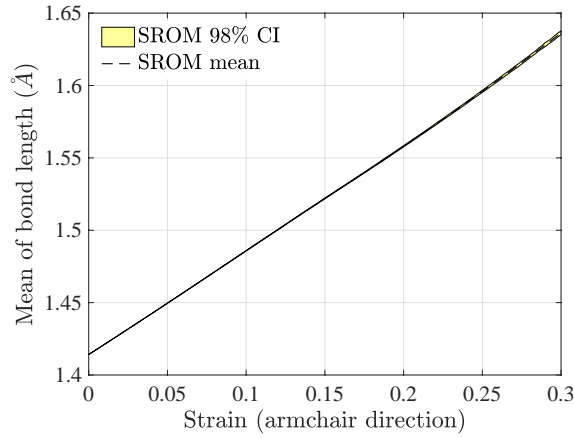


Figure 21: Evolution of the confidence interval for the mean bond length (with SROB  $[\mathbf{W}_0]$ ), as a function of applied engineering strain (armchair direction).

## 5. How to use the proposed methodology to analyze uncertainties in molecular dynamics?

The proposed methodology can be summarized as follows.

1. Define a family of  $N_V$  appropriate interatomic potentials that are adapted to a given quantity of interest. This family contains a reference potential  $V_{\ell^*}$  that will be subsequently used by the SROM.
2. Perform a set of molecular dynamics simulations, potentially with different loading conditions, for all the  $N_V$  selected potentials. Concatenate all the snapshots thus obtained in a deterministic matrix  $[X_0]$  (see above Eq. (10)).
3. Construct the ROB  $[W_0]$ , using a singular value decomposition of  $[X_0]$ . Perform a convergence analyze to determine the number of mode  $n_0$ , for a selected value of  $\varepsilon$ .
4. Define an observable and a suitable cost function. Identify the hyperparameters of the SROB  $[\mathbf{W}_0]$ , using  $V_{\ell^*}$  and following the methodology detailed in Section 3.2.3.

5. For the SROM identified in Step 4, with  $[\mathbf{W}_0]$  and  $V_{\ell^*}$ , perform the propagation of uncertainties and conduct a statistical analysis for any quantity of interest (mean, variance, confidence region, probability distribution).

It should be noticed that the first step is fundamental, since the selected family of potentials generates the level of uncertainties the SROM aims to model.

## 6. Conclusion

A methodology that allows for the integration of model-form uncertainties in molecular dynamics simulations was proposed. The strategy relies on the proper randomization of a reduced-order basis, obtained by the method of snapshots in the configuration space. The relevance of the approach was finally assessed on a variety of situations, including the representation of uncertainties raised by model reduction and interatomic potential selection. The presented framework can be used, in particular, to perform robust MD-based design optimization of materials.

It should be noticed that the proposed approach allows for a complete probabilistic modeling of uncertainties and their propagation for any relevant quantity of interest. This is in contrast with an interval method where only minimum and maximum deterministic bounds, associated with two potentials in the selected family, would be considered. In addition to model-form uncertainties, the SROM also enables the robust, simultaneous treatment of parametric uncertainties on a set of potentials.

## References

- [1] T. Belytschko, S. Xiao, G. Schatz, R. Ruoff, Atomistic simulations of nanotube fracture, *Physical Review B* 65 (2002) 235430.
- [2] J. Tersoff, Modeling solid-state chemistry: Interatomic potentials for multicomponent systems, *Physical Review B* 39 (1989) 5566.
- [3] S. J. Stuart, A. B. Tutein, J. A. Harrison, A reactive potential for hydrocarbons with intermolecular interactions, *The Journal of chemical physics* 112 (2000) 6472–6486.
- [4] J. Los, A. Fasolino, Intrinsic long-range bond-order potential for carbon: Performance in monte carlo simulations of graphitization, *Physical Review B* 68 (2003) 024107.
- [5] J. Tersoff, Empirical interatomic potential for carbon, with applications to amorphous carbon, *Physical Review Letters* 61 (1988) 2879.
- [6] J. Tersoff, Carbon defects and defect reactions in silicon, *Physical review letters* 64 (1990) 1757.
- [7] J. Tersoff, Chemical order in amorphous silicon carbide, *Physical Review B* 49 (1994) 16349.
- [8] M. Tang, S. Yip, Atomistic simulation of thermomechanical properties of  $\beta$ -sic, *Physical Review B* 52 (1995) 15150.
- [9] K. Nordlund, J. Keinonen, T. Mattila, Formation of ion irradiation induced small-scale defects on graphite surfaces, *Physical review letters* 77 (1996) 699.
- [10] F. de Brito Mota, J. Justo, A. Fazzio, Structural properties of amorphous silicon nitride, *Physical Review B* 58 (1998) 8323.
- [11] J. Titantah, D. Lamoen, sp3/sp2 characterization of carbon materials from first-principles calculations: X-ray photoelectron versus high energy electron energy-loss spectroscopy techniques, *Carbon* 43 (2005) 1311–1316.
- [12] L. Lindsay, D. Broido, Optimized tersoff and brenner empirical potential parameters for lattice dynamics and phonon thermal transport in carbon nanotubes and graphene, *Physical Review B* 81 (2010) 205441.
- [13] P. Angelikopoulos, C. Papadimitriou, P. Koumoutsakos, Bayesian uncertainty quantification and propagation in molecular dynamics simulations: a high performance computing framework, *The Journal of chemical physics* 137 (2012) 144103.
- [14] C. Kim, O. Borodin, G. E. Karniadakis, Quantification of sampling uncertainty for molecular dynamics simulation: Time-dependent diffusion coefficient in simple fluids, *Journal of Computational Physics* 302 (2015) 485–508.
- [15] S. Liu, A. Gerisch, M. Rahimi, J. Lang, M. C. Böhm, F. Müller-Plathe, Robustness of a new molecular dynamics-finite element coupling approach for soft matter systems analyzed by uncertainty quantification, *The Journal of Chemical Physics* 142 (2015) 104105.
- [16] T. Le, J. Guilleminot, C. Soize, Stochastic continuum modeling of random interphases from atomistic simulations. Application to a polymer nanocomposite, *Computer Methods in Applied Mechanics and Engineering* 303 (2016) 430 – 449.
- [17] K. Farrell, J. T. Oden, D. Faghihi, A bayesian framework for adaptive selection, calibration, and validation of coarse-grained models of atomistic systems, *Journal of Computational Physics* 295 (2015) 189 – 208.
- [18] K. Farrell-Maupin, J. T. Oden, Adaptive selection and validation of models of complex systems in the presence of uncertainty, *Research in the Mathematical Sciences* 4 (2017) 14.
- [19] R. Ghanem, D. Higdon, H. Owhadi, *Handbook of Uncertainty Quantification*, Springer Cham, 2017.

- [20] F. Rizzi, R. Jones, B. Debusschere, O. Knio, Uncertainty quantification in MD simulations of concentration driven ionic flow through a silica nanopore. ii. uncertain potential parameters, *The Journal of chemical physics* 138 (2013) 194105.
- [21] F. Rizzi, H. N. Najm, B. J. Debusschere, K. Sargsyan, M. Salloum, H. Adalsteinsson, O. M. Knio, Uncertainty quantification in MD simulations. part ii: Bayesian inference of force-field parameters, *Multiscale Modeling & Simulation* 10 (2012) 1460–1492.
- [22] S. T. Reeve, A. Strachan, Error correction in multi-fidelity molecular dynamics simulations using functional uncertainty quantification, *Journal of Computational Physics* 334 (2017) 207–220.
- [23] C. Soize, C. Farhat, A nonparametric probabilistic approach for quantifying uncertainties in low-dimensional and high-dimensional nonlinear models, *International Journal for Numerical Methods in Engineering* 109 (2017) 837–888.
- [24] S. Plimpton, Fast parallel algorithms for short-range molecular dynamics, *Journal of Computational Physics* 117 (1995) 1 – 19.
- [25] C.-H. Lee, J.-S. Chen, Proper orthogonal decomposition-based model order reduction via radial basis functions for molecular dynamics systems, *International Journal for Numerical Methods in Engineering* 96 (2013) 599–627.
- [26] C. de Tomas, I. Suarez-Martinez, N. A. Marks, Graphitization of amorphous carbons: A comparative study of interatomic potentials, *Carbon* 109 (2016) 681–693.
- [27] L. Sirovich, Turbulence and the dynamics of coherent structures. i. coherent structures, *Quarterly of applied mathematics* 45 (1987) 561–571.
- [28] G. Kerschen, J.-c. Golinval, A. F. Vakakis, L. A. Bergman, The method of proper orthogonal decomposition for dynamical characterization and order reduction of mechanical systems: an overview, *Nonlinear dynamics* 41 (2005) 147–169.
- [29] K. Hoang, Y. Fu, J. Song, An hp-proper orthogonal decomposition–moving least squares approach for molecular dynamics simulation, *Computer Methods in Applied Mechanics and Engineering* 298 (2016) 548–575.
- [30] C. Soize, C. Farhat, Probabilistic learning for model-form uncertainties in nonlinear computational mechanics, *International Journal for Numerical Methods in Engineering* online (2018) 1–25.
- [31] X. Zhou, D. K. Ward, M. E. Foster, An analytical bond-order potential for carbon, *Journal of computational chemistry* 36 (2015) 1719–1735.
- [32] D. W. Brenner, O. A. Shenderova, J. A. Harrison, S. J. Stuart, B. Ni, S. B. Sinnott, A second-generation reactive empirical bond order (rebo) potential energy expression for hydrocarbons, *Journal of Physics: Condensed Matter* 14 (2002) 783.
- [33] E. Cadelano, P. L. Palla, S. Giordano, L. Colombo, Nonlinear elasticity of monolayer graphene, *Physical review letters* 102 (2009) 235502.
- [34] C. Lee, X. Wei, J. W. Kysar, J. Hone, Measurement of the elastic properties and intrinsic strength of monolayer graphene, *science* 321 (2008) 385–388.
- [35] J. D. Bernal, The structure of graphite, *Proc. R. Soc. Lond. A* 106 (1924) 749–773.