

# Reliability assessment for large-scale molecular dynamics approximations

Francesca Grogan, Michael Holst, Lee Lindblom, and Rommie Amaro

Citation: *The Journal of Chemical Physics* **147**, 234106 (2017);

View online: <https://doi.org/10.1063/1.5009431>

View Table of Contents: <http://aip.scitation.org/toc/jcp/147/23>

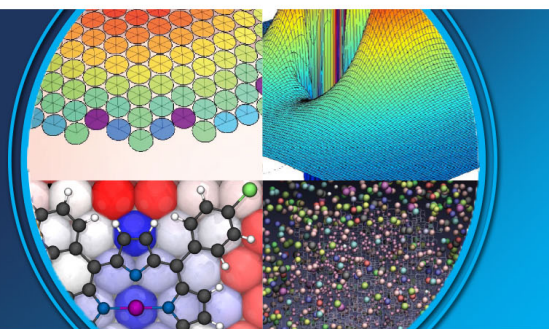
Published by the *American Institute of Physics*

---

---

**AIP** | The Journal of  
Chemical Physics

**PERSPECTIVES**



# Reliability assessment for large-scale molecular dynamics approximations

Francesca Grogan,<sup>1,a)</sup> Michael Holst,<sup>1,2,b)</sup> Lee Lindblom,<sup>1,2,c)</sup> and Rommie Amaro<sup>3,d)</sup>

<sup>1</sup>Center for Computational Mathematics, Department of Mathematics, University of California at San Diego, La Jolla, California 92093, USA

<sup>2</sup>Center for Astrophysics and Space Sciences, University of California at San Diego, La Jolla, California 92093, USA

<sup>3</sup>Department of Chemistry and Biochemistry, University of California at San Diego, La Jolla, California 92093, USA

(Received 16 October 2017; accepted 4 December 2017; published online 20 December 2017)

Molecular dynamics (MD) simulations are used in biochemistry, physics, and other fields to study the motions, thermodynamic properties, and the interactions between molecules. Computational limitations and the complexity of these problems, however, create the need for approximations to the standard MD methods and for uncertainty quantification and reliability assessment of those approximations. In this paper, we exploit the intrinsic two-scale nature of MD to construct a class of large-scale dynamics approximations. The reliability of these methods is evaluated here by measuring the differences between full, classical MD simulations and those based on these large-scale approximations. Molecular dynamics evolutions are non-linear and chaotic, so the complete details of molecular evolutions cannot be accurately predicted even using full, classical MD simulations. This paper provides numerical results that demonstrate the existence of computationally efficient large-scale MD approximations which accurately model certain large-scale properties of the molecules: the energy, the linear and angular momenta, and other macroscopic features of molecular motions. *Published by AIP Publishing.* <https://doi.org/10.1063/1.5009431>

## I. INTRODUCTION

Traditional molecular dynamics (MD) simulations use Newton's classical equations of motion, with an effective potential that models the interactions between atoms, to describe the evolution of molecules.<sup>1–3</sup> This standard MD method has been applied to a variety of problems in biochemistry and condensed matter physics in recent years.<sup>4–11</sup>

The results obtained from these MD simulations can be characterized, qualitatively, as the evolution of the overall position and orientation of each molecule, plus vibrations of the individual atoms about their average positions within each molecule. The time scales associated with the macroscopic bulk motions of the molecules are typically much longer than the time scales associated with the internal atomic vibrations. However, these very short time scale vibrations determine the maximum time steps allowed for accurate solutions of the classical MD evolution equations using standard numerical methods. This fundamental fact, together with the need to simulate extremely large complex molecules in modern biochemical research, means the computational cost of performing MD simulations can be prohibitively large. Computational cost is therefore one of the factors that drives the need to develop approximation methods capable of obtaining reliable simulations of those aspects of the molecular systems of primary interest to researchers.

Another factor that motivates the development of approximation methods for MD simulations is the well-known fact that typical  $n$ -body systems exhibit chaotic behavior in which exponentially divergent evolutions result from small perturbations of initial conditions.<sup>12</sup> It is simply impossible, and therefore pointless to attempt, to simulate all the details in the evolutions of complex molecular systems. Many large-scale properties of such systems, including statistical time averages of various properties, are nevertheless well defined<sup>2,13,14</sup> and these properties are therefore in principle observable and simulatable. These macroscopic characteristics include the positions, average velocities, and other time averages of thermodynamic quantities.

Given these two motivating factors—the need for greater computational efficiency and the fundamental inability to simulate chaotic dynamics in complete microscopic detail—we have developed a new class of large scale MD approximations. We construct these approximations by starting with a new representation of the exact traditional MD equations of motion. The standard MD equations generally use the Cartesian coordinates of the location of each atom as the primary dynamical variables. We transform these Cartesian variables into a new representation that separates the macroscopic location and orientation of each molecule, from the internal degrees of freedom that represent the molecular vibrations. This transformation is (loosely) motivated by Wilson's representation of the internal degrees of freedom of a molecule by normal modes.<sup>15</sup> Our approach differs from Wilson's, however, by providing an exact representation of the molecular motions even when the mode amplitudes are not small. We test this new mode-basis description of MD by

<sup>a)</sup>Electronic mail: fgrogan@ucsd.edu

<sup>b)</sup>Electronic mail: mholst@ucsd.edu

<sup>c)</sup>Electronic mail: llindblom@ucsd.edu

<sup>d)</sup>Electronic mail: ramaro@ucsd.edu

comparing the results of numerical simulations using it with those based on the standard classical MD equations. The results of these numerical tests, described in more detail in Sec. III B, confirm that our new mode-basis representation of MD is exact.

In our mode-basis representation of MD, the large-scale degrees of freedom are cleanly separated from the collection of mode amplitudes that describe the internal vibration degrees of freedom of the molecule. It is straightforward to replace the exact equations that determine the evolution of these mode amplitudes, with various approximate expressions. For example, the mode amplitudes could be chosen to satisfy the analytic sinusoidal-in-time expressions derived from the small-amplitude normal-mode equations. Another possibility would be to use the normal-mode sinusoidal-in-time expressions for some number of the highest frequency modes, while evolving the remaining lower frequency mode amplitudes numerically using the exact evolution equations. This approach would be qualitatively similar to that used in the constraint algorithms like SHAKE and RATTLE.<sup>16,17</sup> Another possibility would simply be to ignore the internal degrees of freedom of the atoms completely by setting the mode amplitudes equal to zero, i.e., to their expected long time scale averages. We present numerical tests that compare different approximations of this type with the results of the exact MD evolutions. The results of these numerical tests are described in more detail in Sec. III C.

We assess the reliability of our new large-scale approximation methods by applying the techniques of uncertainty quantification (UQ).<sup>18–22</sup> Here we focus attention on measuring the accuracy of our new large-scale approximation methods by comparing them to standard classical MD, setting aside other important issues such as time-stepper integration errors, errors in the molecular interaction potential model, errors that arise from the use of classical rather than a fully quantum description of MD, etc. The MD evolution equations are highly non-linear, as are the equations for our large-scale approximations. It is possible to derive rigorous analytic mathematical bounds on the errors in our large-scale approximation methods.<sup>23</sup> However the bounds we have obtained in this way are quite weak, and do not provide a good estimate of the size of the actual large-scale approximation errors in practical simulations. Therefore we focus the discussion of our UQ analysis in this paper on making detailed numerical comparisons between full classical MD simulations and those obtained for identical molecular systems using our new large-scale approximations.

Since MD simulations are typically performed to estimate the values of various macroscopic observables of the molecules, we have focused our uncertainty quantification analysis on assessing the errors in the large-scale approximation values of those quantities. In particular we evaluate the errors in the energy, the linear and angular momenta, and the errors in the positions and orientations of each molecule. Our numerical results show that all the large-scale approximations tested here are linear momentum conserving (MC), and consequently the positions and velocities of the molecular centers of mass are also determined exactly. Some of our large-scale approximations tested here also conserve energy and

angular momentum exactly. Angular momentum conservation does not, however, guarantee that the orientations or angular velocities of the molecules are determined accurately. We show that these orientation features evolve chaotically in MD systems, and are therefore unpredictable even in full classical MD simulations.

The remainder of this paper is organized as follows. In Sec. II, we derive a new mode-basis representation of the molecular dynamics evolution equations, and then use them to derive several large-scale molecular dynamics approximations. We have implemented these equations in a numerical MD evolution code and have used this code to evolve simple models of several of the smaller fullerene molecules:  $C_{20}$ ,  $C_{26}$ ,  $C_{60}$  and  $C_{70}$ . In Sec. III we discuss the results of our numerical simulations of these molecules using both standard classical MD and several large-scale approximations, highlighting the uncertainty quantification of the macroscopic properties of these molecules. We conclude by summarizing and discussing our results briefly in Sec. IV.

## II. CLASSICAL MOLECULAR DYNAMICS

Classical molecular dynamics (MD) uses Newton's equations to describe the motions of the atoms (represented as point particles) that make up the molecules being studied. We use the notation  $\vec{x}_A(t)$  to denote the Cartesian coordinates of the location of atom  $A$  as a function of time. The classical MD equations of motion for the atoms that make up a molecule (or collection of molecules) are therefore given by

$$m_A \frac{d^2 \vec{x}_A}{dt^2} = - \frac{\partial U(\vec{x}_B)}{\partial \vec{x}_A}, \quad (1)$$

where  $m_A$  is the mass of atom  $A$  and  $U(\vec{x}_B)$  is the effective potential energy function that describes the interactions between the atoms. This potential energy will in general be a non-linear function of the locations of all of the atoms. Consequently the MD evolution equations, Eq. (1), are non-linear and strongly coupled. The solutions to these equations therefore display the typical characteristics of chaotic  $n$ -body systems,<sup>2,13,14</sup> i.e., while most details of the evolution of a particular initial molecular state cannot be predicted but certain "statistical" features of the evolution can.

In this section, we develop a new class of large-scale approximations to the classical MD equations of motion, Eq. (1). These new approximations are designed to provide a more efficient way to evaluate some of the observable "statistical" or "thermodynamic" features of MD systems. We construct these approximations in two steps. In the first step, in Sec. II A, we transform the exact classical MD equations into a representation that cleanly separates some of the observable macroscopic degrees of freedom of a molecule from the internal degrees of freedom that determine its microscopic state. We refer to this new representation of the MD equations as a mode-basis representation because the choice of variables used to describe the microscopic state of a molecule is based (loosely) on the normal-mode description of molecular vibrations.<sup>15,24</sup> The transformation we use to construct this mode-basis representation is exact; however, so it is simply a change of variables

for the classical MD system given in Eq. (1). We present the exact MD equations of motion for the mode-basis representation in Sec. II B. The second step in the development of the new class of MD approximations is to replace the exact mode-amplitude evolution equations with various approximations. We present several examples in Sec. II C that use simple analytic expressions to approximate the evolutions of some or all of the mode amplitudes which describe the internal molecular degrees of freedom.

### A. Mode-basis representation of MD

The mode-basis representation of MD is obtained by transforming the Cartesian coordinate variables,  $\vec{x}_A$ , used in the standard representation into new variables that separate into (i) a set that describes the macroscopic location and orientation of the molecule and (ii) another set that describes the molecules' microscopic vibrational dynamics. The macroscopic location of a molecule is represented by its center of mass,  $\vec{x}_{CM}(t)$ , defined by

$$\vec{x}_{CM}(t) = \frac{1}{M} \sum_A m_A \vec{x}_A(t), \quad (2)$$

where  $M = \sum_A m_A$  is the total mass of the molecule.<sup>28</sup> The macroscopic orientation of a molecule is represented by a time-dependent rotation matrix  $\mathbf{R}(t)$ . This matrix provides the transformation between a reference frame fixed to and co-moving with the molecule, and the global inertial frame used to describe the atoms in the standard representation of MD. We use the notation  $\vec{x}_{0A} + \delta\vec{x}_A(t)$  to denote the location of atom  $A$  in the molecule's co-moving reference frame, where  $\vec{x}_{0A}$  represents the time independent equilibrium position, and  $\delta\vec{x}_A(t)$  represents the displacement from equilibrium (which is not assumed to be small) of this atom. The global Cartesian coordinate location of atom  $A$  is determined by these macroscopic variables—the center of mass,  $\vec{x}_{CM}(t)$  and the orientation matrix,  $\mathbf{R}(t)$ —along with the internal dynamical variables  $\delta\vec{x}_A$ ,

$$\vec{x}_A(t) = \vec{x}_{CM}(t) + \mathbf{R}(t) \cdot [\vec{x}_{0A} + \delta\vec{x}_A(t)]. \quad (3)$$

In the mode-basis representation of MD, the internal microscopic degrees of freedom of a molecule are described by the variables  $\delta\vec{x}_A$ . Unfortunately these variables are not independent, so special care must be taken to isolate the truly independent degrees of freedom they represent. To see this more clearly, note that in the standard MD representation, there are  $3N$  variables,  $\vec{x}_A$ , needed to represent the configuration state of a molecule having  $N$  atoms. The macroscopic variables introduced above,  $\{\vec{x}_{CM}, \mathbf{R}\}$ , represent 6 of these degrees of freedom (since there is a three-dimensional space of rotation matrices  $\mathbf{R}$ ). Consequently there can only be  $3N - 6$  truly independent internal microscopic degrees of freedom among the  $3N$  variables  $\delta\vec{x}_A$ . To isolate these independent degrees of freedom, we introduce a collection of “mode-basis” vectors  $\vec{e}_A^\mu$ , where the index  $\mu$  labels the  $3N - 6$  vectors representing those independent degrees of freedom. Without loss of generality, we can normalize these basis vectors

$$\delta^{\mu\nu} = \sum_A \frac{m_A}{M} \vec{e}_A^\mu \cdot \vec{e}_A^\nu, \quad (4)$$

where  $\delta^{\mu\nu}$  is the Kronecker delta. To ensure that the  $\vec{e}_A^\mu$  are independent from the macroscopic variables,  $\{\vec{x}_{CM}, \mathbf{R}\}$ , we choose them to be orthogonal to any overall translation or rotation of the molecule,

$$0 = \sum_A \frac{m_A}{M} \vec{e}_A^\mu, \quad (5)$$

$$0 = \sum_A \frac{m_A}{M} \vec{e}_A^\mu \times \vec{x}_{0A}. \quad (6)$$

Appendix A explains in detail why Eqs. (5) and (6) are the conditions needed to enforce the translation and rotation invariance of the eigenvectors  $\vec{e}_A^\mu$ . Given any collection of mode-basis vectors satisfying Eqs. (4)–(6), it is straightforward to write down a general expression for  $\delta\vec{x}_A$  in terms of  $3N - 6$  independent mode-amplitude functions  $\mathcal{A}_\mu(t)$ ,

$$\delta\vec{x}_A = \sum_\mu \mathcal{A}_\mu(t) \vec{e}_A^\mu. \quad (7)$$

Using this expression and Eq. (3), it is now possible to write down the transformation between the mode-basis representation variables,  $\{\vec{x}_{CM}, \mathbf{R}, \mathcal{A}_\mu\}$ , and the Cartesian coordinates,  $\vec{x}_A$  used in the standard representation of MD,

$$\vec{x}_A(t) = \vec{x}_{CM}(t) + \mathbf{R}(t) \cdot \left[ \vec{x}_{0A} + \sum_\mu \mathcal{A}_\mu(t) \vec{e}_A^\mu \right]. \quad (8)$$

There are an infinite number of ways to choose the mode-basis vectors,  $\vec{e}_A^\mu$ . One natural choice is to let  $\vec{e}_A^\mu$  be the eigenvectors of the Hessian of the potential energy function,

$$0 = -m_A \omega_\mu^2 \vec{e}_A^\mu + \sum_B \frac{\partial^2 U}{\partial \vec{x}_B \partial \vec{x}_A} \cdot \vec{e}_B^\mu. \quad (9)$$

The Hessian matrix,  $\partial^2 U / \partial \vec{x}_B \partial \vec{x}_A$ , in this equation is to be evaluated at the equilibrium state of the molecule where  $\vec{x}_A = \vec{x}_{0A}$ . Since the Hessian is a symmetric real matrix, the eigenvalues  $\omega_\mu^2$  and eigenvectors  $\vec{e}_A^\mu$  are also real. Equation (9) is equivalent to Newton's equation of motion, Eq. (1), for the case of very small amplitude oscillations about its equilibrium state. This choice of eigenvectors is therefore particularly useful for isolating the individual microscopic degrees of freedom of a molecule. The connection of these eigenvectors to the classical normal-mode analysis of molecular vibrations<sup>15,24</sup> motivated our choice of the name “mode-basis” representation. The eigenvalues  $\omega_\mu^2$  in Eq. (9) are non-negative for any stable molecule, and except for the six zero-frequency modes that correspond to rigid rotations and translations of the molecule, these eigenvalues are strictly positive (generically). Since the Hessian matrix is symmetric, the eigenvectors  $\vec{e}_A^\mu$  form a complete basis for the  $\delta\vec{x}_A$  that satisfy (or in the case of degenerate eigenvalues, can be chosen to satisfy) the orthogonality conditions, Eq. (4). Appendix A shows that they also satisfy the constraints, Eqs. (5) and (6).

The standard MD equations of motion, Eq. (1), are second-order ordinary differential equations. Therefore both

the position  $\vec{x}_A$  and the velocity  $\vec{v}_A = d\vec{x}_A/dt$  of each atom are needed to determine the full dynamical state of a molecule. The analog of these velocity degrees of freedom for the macroscopic variables is the center of mass velocity  $\vec{v}_{CM} = d\vec{x}_{CM}/dt$ , and the time derivative of the orientation matrix  $d\mathbf{R}/dt$ . It is convenient and customary to express the time derivative of the orientation matrix as an angular velocity vector,  $\vec{\Omega}$ , in the following way. The matrix  $*\Omega$  defined by

$$*\Omega = -\frac{d\mathbf{R}}{dt} \cdot \mathbf{R}^{-1} \quad (10)$$

is anti-symmetric for any rotation matrix  $\mathbf{R}$ . Therefore  $*\Omega$  is dual to a vector  $\vec{\Omega}$ ,

$$*\Omega_{ij} = \sum_k \epsilon_{ijk} \Omega^k, \quad (11)$$

where  $\epsilon_{ijk}$  is the totally anti-symmetric tensor with  $\epsilon_{xyz} = 1$  in Cartesian coordinates. The time derivative  $d\mathbf{R}/dt$  is therefore given by

$$\frac{d\mathbf{R}}{dt} = -*\Omega \cdot \mathbf{R} \quad (12)$$

or in component notation,

$$\frac{dR_{ij}}{dt} = -\sum_{k\ell} \epsilon_{ik\ell} \Omega^\ell R_{kj}. \quad (13)$$

Using these results, we can now write down the complete set of transformation equations between the Cartesian coordinate variables,  $\{\vec{x}_A, \vec{v}_A\}$ , used in the standard representations of classical MD, and the variables,  $\{\vec{x}_{CM}, \vec{v}_{CM}, \mathbf{R}, \vec{\Omega}, \mathcal{A}_\mu, d\mathcal{A}_\mu/dt\}$  used in our new mode-basis representation,

$$\vec{x}_A(t) = \vec{x}_{CM}(t) + \Delta\vec{x}_A(t), \quad (14)$$

$$\vec{v}_A(t) = \vec{v}_{CM}(t) + \vec{\Omega}(t) \times \Delta\vec{x}_A(t) + \sum_\mu \frac{d\mathcal{A}_\mu(t)}{dt} \mathbf{R}(t) \cdot \vec{e}_A^\mu, \quad (15)$$

where  $\Delta\vec{x}_A$  is given by

$$\Delta\vec{x}_A(t) = \mathbf{R}(t) \cdot \left[ \vec{x}_{0A} + \sum_\mu \mathcal{A}_\mu(t) \vec{e}_A^\mu \right]. \quad (16)$$

## B. Evolution equations for mode-basis MD

The evolution equations for the mode-basis dynamical variables,  $\{\vec{x}_{CM}, \mathbf{R}, \vec{\Omega}, \mathcal{A}_\mu\}$ , are determined from Eq. (1) using the transformation given in Eq. (8). The resulting equations can be written in the form

$$\frac{d^2\vec{x}_{CM}}{dt^2} = -\sum_A \frac{1}{M} \frac{\partial U}{\partial \vec{x}_A}, \quad (17)$$

$$\frac{d\mathbf{R}}{dt} = -*\Omega \cdot \mathbf{R}, \quad (18)$$

$$\frac{d\vec{\Omega}}{dt} = \vec{\mathcal{U}}^{-1} \cdot \vec{\mathcal{V}}, \quad (19)$$

$$\frac{d^2\mathcal{A}_\mu}{dt^2} = \sum_\nu S^{\mu\nu} \frac{d\mathcal{A}_\nu}{dt} + \sum_\nu \mathcal{T}^{\mu\nu} \mathcal{A}_\nu + \mathcal{F}^\mu. \quad (20)$$

The tensor  $\vec{\mathcal{U}}$  and vector  $\vec{\mathcal{V}}$  that appear in Eq. (19) are functions of  $\{\vec{x}_{CM}, \mathbf{R}, \vec{\Omega}, \mathcal{A}_\mu, d\mathcal{A}_\mu/dt\}$  given by

$$\vec{\mathcal{U}} = \sum_A \left( \frac{m_A}{M} \right) \left[ \mathbf{R}^{-1} (\vec{x}_{0A} + \delta\vec{x}_A) \cdot \vec{x}_{0A} - (\vec{x}_{0A} + \delta\vec{x}_A) \otimes (\mathbf{R} \cdot \vec{x}_{0A}) \right], \quad (21)$$

$$\vec{\mathcal{V}} = \sum_A \left( \frac{m_A}{M} \right) \left[ 2(\vec{x}_{0A} \cdot \mathbf{R}^{-1} \cdot \vec{\Omega}) \frac{d\delta\vec{x}_A}{dt} - 2 \left( \vec{x}_{0A} \cdot \frac{d\delta\vec{x}_A}{dt} \right) (\mathbf{R}^{-1} \cdot \vec{\Omega}) - (\vec{\Omega} \cdot \Delta\vec{x}_A) \vec{x}_{0A} \times (\mathbf{R}^{-1} \cdot \vec{\Omega}) - \frac{1}{m_A} \vec{x}_{0A} \times \left( \mathbf{R}^{-1} \cdot \frac{\partial U}{\partial \vec{x}_A} \right) \right], \quad (22)$$

where  $\delta\vec{x}_A$  is given in Eq. (7). Similarly the quantities  $S^{\mu\nu}$ ,  $\mathcal{T}^{\mu\nu}$ , and  $\mathcal{F}^\mu$  that appear in Eq. (20) are functions of  $\{\vec{x}_{CM}, \mathbf{R}, \vec{\Omega}, \mathcal{A}_\mu, d\mathcal{A}_\mu/dt\}$  given by

$$S^{\mu\nu} = 2 \sum_A \frac{m_A}{M} (\mathbf{R} \cdot \vec{e}_A^\mu) \cdot [(\mathbf{R} \cdot \vec{e}_A^\nu) \times \vec{\Omega}], \quad (23)$$

$$\mathcal{T}^{\mu\nu} = \vec{\Omega} \cdot \vec{\Omega} \delta^{\mu\nu} + \sum_A \frac{m_A}{M} (\mathbf{R} \cdot \vec{e}_A^\mu) \cdot \left[ (\mathbf{R} \cdot \vec{e}_A^\nu) \times \frac{d\vec{\Omega}}{dt} \right] - \sum_A \frac{m_A}{M} [(\mathbf{R} \cdot \vec{e}_A^\mu) \cdot \vec{\Omega}] [(\mathbf{R} \cdot \vec{e}_A^\nu) \cdot \vec{\Omega}], \quad (24)$$

$$\mathcal{F}^\mu = -\sum_A \frac{1}{M} (\mathbf{R} \cdot \vec{e}_A^\mu) \cdot \frac{\partial U(\vec{x}_A)}{\partial \vec{x}_A} + \sum_A \frac{m_A}{M} (\mathbf{R} \cdot \vec{e}_A^\mu) \cdot \left[ (\mathbf{R} \cdot \vec{x}_{0A}) \times \frac{d\vec{\Omega}}{dt} \right] - \sum_A \frac{m_A}{M} \{ [(\mathbf{R} \cdot \vec{e}_A^\mu) \cdot \vec{\Omega}] [(\mathbf{R} \cdot \vec{x}_{0A}) \cdot \vec{\Omega}] - (\vec{e}_A^\mu \cdot \vec{x}_{0A}) (\vec{\Omega} \cdot \vec{\Omega}) \}. \quad (25)$$

The  $d\vec{\Omega}/dt$  that appears in Eqs. (24) and (25) are to be replaced by the expression on the right side of Eq. (19). With those replacements, the expressions on the right sides of Eqs. (17)–(20) depend only on  $\{\vec{x}_{CM}, d\vec{x}_{CM}/dt, \mathbf{R}, \vec{\Omega}, \mathcal{A}_\mu, d\mathcal{A}_\mu/dt\}$ .

While the derivations leading to Eqs. (17)–(20) are straightforward, they are lengthy and have not been reproduced here in detail. Those derivations can be summarized however as follows. Equation (17) is obtained by inserting Eq. (1) into the second time derivative of Eq. (2). Equation (18) follows trivially from Eqs. (10) and (11). The derivation of Eq. (19) is more complicated. It is obtained by projecting Eq. (1) onto the three independent generators of rigid rotations of the molecule, i.e., the vectors  $\hat{\theta} \times \vec{x}_{0A}$  where  $\hat{\theta}$  is a unit vector whose direction determines the axis of rotation. Similarly Eq. (20) is obtained by projecting Eq. (1) onto each of the mode-basis vectors  $\vec{e}_A^\mu$ .

Finally, we note that while Eq. (18) determines the evolution of the rotation matrix  $\mathbf{R}(t)$ , solving this equation numerically directly in this form is problematic. Instead it is better to adopt some parameterization for the rotation matrices, e.g., using Euler angles, and then to solve numerically the differential equations implied by Eq. (18) for those parameters. In our



numerical work, we have adopted the “quaternion” parameterization of rotation matrices, in which each rotation matrix is represented by four parameters  $\{q_0, q_1, q_2, q_3\}$  with  $q_0^2 + q_1^2 + q_2^2 + q_3^2 = 1$ . Appendix B describes this quaternion representation and explicitly gives the representation of Eq. (18) in terms of these parameters. We point out that the version of the quaternion evolution equations used here introduces a new (so far as we know) constraint damping mechanism that ensures the constraint,  $q_0^2 + q_1^2 + q_2^2 + q_3^2 = 1$ , remains satisfied by the numerical solution.

### C. Large-scale MD approximations

Equations (17)–(20) are a well-posed system of ordinary differential equations that give an exact representation of classical MD in terms of the mode-basis dynamical variables  $\{\vec{x}_{CM}, \mathbf{R}, \vec{\Omega}, \mathcal{A}_\mu\}$ . The idea of our large-scale MD approximations is to use some subset of the exact equations to determine the macroscopic degrees of freedom,  $\{\vec{x}_{CM}, \mathbf{R}, \vec{\Omega}\}$ , while using simpler approximate equations to determine the evolution of the internal vibrational degrees of freedom  $\mathcal{A}_\mu$ .

The most straightforward way to construct a large-scale approximation uses Eqs. (17)–(19) to determine  $\{\vec{x}_{CM}, \mathbf{R}, \vec{\Omega}\}$ , while replacing Eq. (20) with some approximate equation for  $\mathcal{A}_\mu$ . Perhaps the most natural approximation for  $\mathcal{A}_\mu$ , which we refer to as the sinusoidal mode amplitude (SMA) approximation, would be to set the mode amplitudes  $\mathcal{A}_\mu(t)$  to their small-amplitude perturbation solution values:

$$\mathcal{A}_\mu(t) = \mathcal{A}_\mu^0 \sin(\omega_\mu t + \varphi_\mu), \quad (26)$$

where  $\omega_\mu$  is the mode frequency determined from Eq. (9), while  $\mathcal{A}_\mu^0$  and  $\varphi_\mu$  are constants that specify the amplitude of phase of each mode. In this approximation, Eq. (26) replaces Eq. (20) and is used to evaluate the right sides of Eqs. (17)–(19). Those equations for the macroscopic degrees of freedom  $\{\vec{x}_{CM}, \mathbf{R}, \vec{\Omega}\}$  are then solved numerically. The system of equations being solved numerically is therefore reduced from  $6N$  first-order equations for the exact MD system, to just twelve for this large-scale approximation. The use of this approximation eliminates the need to evaluate the complicated quantities  $\mathcal{S}^{\mu\nu}$ ,  $\mathcal{T}^{\mu\nu}$  and  $\mathcal{F}^\mu$  that appear on the right side of Eq. (20) numerically. This reduction in the number of equations to be solved numerically, as well as the reduction in the need to evaluate the complicated expressions that occur on the right side of Eq. (20) considerably reduces the computational cost of implementing the SMA approximation.

Another plausible approximation, which we refer to as the zero mode amplitude (ZMA) approximation, simply sets all the mode amplitudes to zero:

$$\mathcal{A}_\mu(t) = 0. \quad (27)$$

We expect the long time averages of the positions of the atoms to be their equilibrium positions  $\vec{x}_{0A}$ . Thus we expect the time averages of the mode amplitudes  $\mathcal{A}_\mu(t)$  to be zero, as they are for example in the SMA approximation given in Eq. (26). This should reduce the computational cost of evaluating the right sides of Eqs. (17)–(19) in the ZMA approximation even below those costs in the SMA approximation. In addition the ZMA approximation eliminates all the short time scale

effects associated with the molecular vibrations, so it should be possible to use much larger time steps to determine the macroscopic variables  $\{\vec{x}_{CM}, \mathbf{R}, \vec{\Omega}\}$  in this approximation, and thus to reduce the computational cost even below those for the SMA approximation.

We note that hybrid approximations can easily be constructed as well. In these hybrid approximations some of the mode amplitudes are set to the SMA or the ZMA approximations given in Eq. (26) or (27), while the remaining amplitudes are determined numerically by solving Eq. (20). This approach might be appropriate for systems having a few modes with oscillation time scales comparable to the time scales associated with the macroscopic properties of the molecule. In such cases those low frequency modes could be treated exactly while the approximations could still be used for the majority of modes having much shorter oscillation time scales.

Somewhat more sophisticated large-scale approximations can also be obtained by choosing the equations of motion for the macroscopic variables,  $\{\vec{x}_{CM}, \mathbf{R}, \vec{\Omega}\}$ , from the particular combination of the exact equations that determine the evolution of the macroscopic linear and angular momentum,  $\vec{P}$  and  $\vec{J}$ , of each molecule. These quantities are defined by

$$\vec{P} = \sum_A m_A \frac{d\vec{x}_A}{dt}, \quad (28)$$

$$\vec{J} = \sum_A m_A \Delta\vec{x}_A \times \frac{d\Delta\vec{x}_A}{dt}, \quad (29)$$

where  $\Delta\vec{x}_A$ , defined in Eq. (14), is the position of each atom relative to the center of mass of the molecule.<sup>29</sup> The time derivatives of these quantities can be written in terms of the macroscopic variables as

$$\frac{d\vec{P}}{dt} = M \frac{d^2\vec{x}_{CM}}{dt^2}, \quad (30)$$

$$\frac{d\vec{J}}{dt} = \tilde{\mathcal{J}}_\Omega \cdot \frac{d\vec{\Omega}}{dt} + \vec{\mathcal{J}}, \quad (31)$$

where  $\tilde{\mathcal{J}}_\Omega$  and  $\vec{\mathcal{J}}$  are given by

$$\tilde{\mathcal{J}}_\Omega = \sum_A m_A (\Delta\vec{x}_A \cdot \Delta\vec{x}_A \mathbf{I} - \Delta\vec{x}_A \otimes \Delta\vec{x}_A), \quad (32)$$

$$\vec{\mathcal{J}} = \sum_A m_A \Delta\vec{x}_A \times \vec{B}_A, \quad (33)$$

and where  $\vec{B}_A$  is given by

$$\begin{aligned} \vec{B}_A = \sum_\mu \left( \frac{d^2\mathcal{A}_\mu}{dt^2} + 2 \frac{d\mathcal{A}_\mu}{dt} \vec{\Omega} \times \right) \mathbf{R} \cdot \vec{e}_A^\mu \\ + (\vec{\Omega} \otimes \vec{\Omega} - \vec{\Omega} \cdot \vec{\Omega} \mathbf{I}) \cdot \Delta\vec{x}_A. \end{aligned} \quad (34)$$

If we assume the mode amplitudes  $\mathcal{A}_\mu(t)$  are predetermined by some approximate expressions, like those in Eq. (26) or (27) for example, then  $\tilde{\mathcal{J}}_\Omega$  and  $\vec{\mathcal{J}}$  depend only on the large scale variables  $d\vec{x}_{CM}/dt$ ,  $\mathbf{R}$ ,  $\vec{\Omega}$ , but not on their time derivatives. Equations (30) and (31) can therefore be used to construct an alternate set of approximate evolution equations for the large scale variables. In particular an evolution equation for the center-of-mass motion of each molecule

can be obtained by setting the rate of change of the total momentum equal to the total external force acting on the molecule:

$$\frac{d\vec{P}}{dt} = - \sum_A \frac{\partial U}{\partial \vec{x}_A} = M \frac{d^2 \vec{x}_{CM}}{dt^2}. \quad (35)$$

Similarly an equation for  $d\vec{\Omega}/dt$  can be obtained by setting  $d\vec{J}/dt$  equal to the total external torque acting on the molecule:

$$\frac{d\vec{J}}{dt} = - \sum_A \Delta \vec{x}_A \times \frac{\partial U}{\partial \vec{x}_A} = \vec{\mathcal{J}}_{\Omega} \cdot \frac{d\vec{\Omega}}{dt} + \vec{\mathcal{J}}. \quad (36)$$

The resulting evolution equations for  $\vec{x}_{CM}$  and  $\vec{\Omega}$  are given by

$$\frac{d^2 \vec{x}_{CM}}{dt^2} = - \frac{1}{M} \sum_A \frac{\partial U}{\partial \vec{x}_A}, \quad (37)$$

$$\frac{d\vec{\Omega}}{dt} = - (\vec{\mathcal{J}}_{\Omega})^{-1} \cdot \left[ \vec{\mathcal{J}} + \sum_A \Delta \vec{x}_A \times \frac{\partial U}{\partial \vec{x}_A} \right]. \quad (38)$$

Equations (37) and (38) represent somewhat different projections of the exact MD equations, Eq. (1), than those given in Eqs. (17) and (19). Therefore Eqs. (37) and (38) together with Eq. (18), provide an alternate somewhat different set of evolution equations for the macroscopic variables  $\{\vec{x}_{CM}, \mathbf{R}, \vec{\Omega}\}$ . We refer to these alternate equations as the momentum conserving (MC) large-scale approximation. These equations can be solved using any predetermined approximate form for the mode amplitudes  $\mathcal{A}_{\mu}(t)$ . In this paper we explore the two possibilities discussed above: We refer to the momentum conserving approximation using sinusoidal mode amplitudes, Eq. (26), as the MCSMA approximation and the momentum conserving approximation using zero mode amplitudes, Eq. (27), as the MCZMA approximation.

### III. RELIABILITY TESTING

In this section, we assess the reliability of the large-scale MD approximations introduced in Sec. II C. We do this by comparing the values of the macroscopic variables,  $\{\vec{x}_{CM}, \vec{v}_{CM}, \mathbf{R}, \vec{\Omega}\}$ , computed numerically using the exact MD equations, with those computed using several examples of large-scale MD approximations. We also compare how well these various methods conserve the total energy  $E$ , the total momentum  $\vec{P}$ , and the total angular momentum  $\vec{J}$  of each

molecule. The remainder of this section is organized as follows. Section III A describes in detail the model problem and the numerical methods used to solve the MD equations for these tests. Section III B presents the results of numerical tests that confirm that the mode-basis representation of the exact MD equations, introduced in Sec. II B, gives the same results as the standard Cartesian representation for this model problem. Section III C gives the results of our numerical tests for the large-scale approximations SMA, ZMA, MCSMA, and MCZMA introduced in Sec. II C. Finally in Sec. III D, we compare the computational efficiency of these various methods when applied to our model problem.

#### A. Model problem

We use a collection of simple molecules, the fullerenes  $C_{20}$ ,  $C_{26}$ ,  $C_{60}$ , and  $C_{70}$ , to study the reliability of the large-scale MD approximations introduced in Sec. II C. These molecules consist entirely of trivalent carbon atoms located at the vertices of convex polyhedra. Figure 1 illustrates the topological bond connections between the atoms (but not the actual geometrical shapes) of the molecules used in our tests.

We use a simplified version of the CHARMM27 model for the potential energy function  $U(\vec{x}_A)$  that determines the interactions between the atoms. This interaction potential is given by

$$U = \frac{1}{2} \kappa_b \sum_{\text{bonds } (CD)} (r_{CD} - L_b)^2 + \frac{1}{2} \kappa_{\theta} \sum_{\text{angles } (CDE)} (\theta_{CDE} - \theta_b)^2, \quad (39)$$

where  $L_b$  and  $\theta_b$  are the lengths and angles of equilibrium molecular bonds in this simple model, and where  $r_{CD}$  and  $\theta_{CDE}$  represent, respectively, the distance between atoms  $C$  and  $D$  and the angle formed by the bonds between atom  $D$  with atoms  $C$  and  $E$ ,

$$r_{CD}^2 = (\vec{x}_C - \vec{x}_D) \cdot (\vec{x}_C - \vec{x}_D), \quad (40)$$

$$\cos \theta_{CDE} = \frac{(\vec{x}_C - \vec{x}_D) \cdot (\vec{x}_E - \vec{x}_D)}{r_{CD} r_{DE}}. \quad (41)$$

The sums in Eq. (39) are taken, respectively, over the collection of bonds ( $CD$ ) between the pairs of atoms  $C$  and  $D$  and over the collection of angles ( $CDE$ ) formed by the bond between atoms  $C$  and  $D$  and the bond between atoms  $D$  and  $E$ . We use the following values for the carbon-carbon bond parameters:  $\kappa_b = 305 \text{ kcal}/(\text{\AA}^2/\text{mol})$ ,  $L_b = 1.375 \text{ \AA}$ , and  $\theta_b = 120^\circ$  taken from the CHARMM27 force field

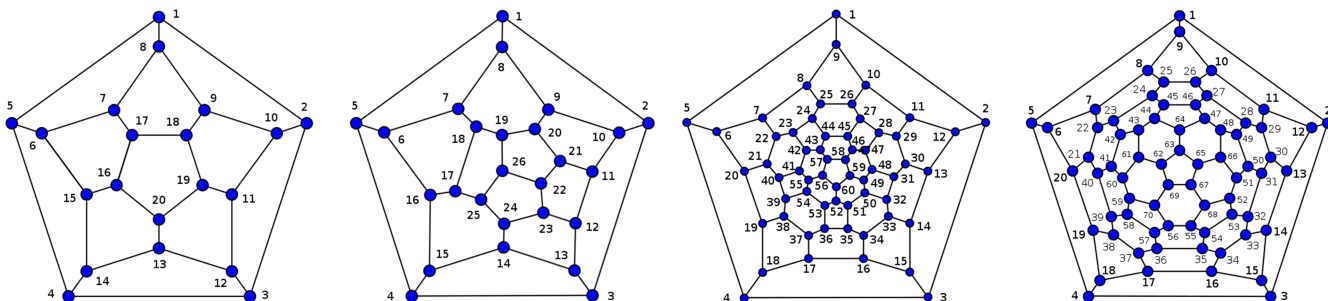


FIG. 1. Figures show the topological bond connections for the  $C_{20}$ ,  $C_{26}$ ,  $C_{60}$ , and  $C_{70}$  molecules used in our test MD simulations. Each atom is labeled with a number that represents the value of the index  $A$  for that atom (adapted from figures produced by R.A. Nonnenmacher and published in Wikipedia at <https://en.wikipedia.org/wiki/Fullerene> under the Creative Commons Attribution-Share Alike 3.0 Unported license).

parameters for these carbon-carbon bonds.<sup>25</sup> For simplicity in coding up our tests, we left out the standard torsion-angle bond interactions from the potential energy. Without those torsion-angle forces, the fullerene molecules are unstable with our simplified interaction potential when the CHARMM27 value is used for the bond angle force constant,  $\kappa_\theta = 40$  kcal/(rad<sup>2</sup>/mol). Therefore, we have increased the value of  $\kappa_\theta$  used in our tests to  $\kappa_\theta = 305$  kcal/(rad<sup>2</sup>/mol) to achieve stability. Our purpose here is to test the robustness of our large-scale approximations. These approximations should succeed or fail independent of the details of the interaction potential model being used, so we do not think it matters that our simplified potential model is not state of the art.

The first step in our analysis of these molecules is to determine their equilibrium configurations,  $\vec{x}_{0A}$ , for the interaction potential  $U(\vec{x}_A)$  given in Eq. (39). We do this by finding the energy minimum where  $\partial U/\partial \vec{x}_A = 0$ . We use the Fletcher-Reeves-Polak-Ribiere version of the conjugate gradient method with line minimizations to find these equilibrium states,  $\vec{x}_{0A}$ , numerically.<sup>26</sup> Given an equilibrium state, we next evaluate the Hessian matrix  $\partial^2 U/\partial \vec{x}_A \partial \vec{x}_B$  numerically for that state and solve Eq. (9) to determine the eigenvalues  $\omega_\mu$  and eigenvectors  $\vec{e}_A^\mu$ . We use Householder reduction to transform  $\partial^2 U/\partial \vec{x}_A \partial \vec{x}_B$  to tridiagonal form, followed by a traditional QL algorithm with implicit shifts to determine the eigenvalues and eigenvectors numerically.<sup>26</sup> These eigenvectors are then projected and normalized so they satisfy Eqs. (4)–(6) to double precision accuracy numerically.

We construct initial data for our test evolutions by choosing values for the mode-basis variables  $\{\vec{x}_{CM}, \vec{v}_{CM}, \mathbf{R}, \vec{\Omega}, \mathcal{A}_\mu, d\mathcal{A}_\mu/dt\}$  that are appropriate for a thermodynamic equilibrium state at temperature  $T$ . Following the equipartition theorem, we fix the values for each of the mode-basis variables so that each degree of freedom of the molecule has energy  $\frac{1}{2}kT$ , where  $k = 1.9872 \times 10^{-3}$  kcal/(mol K) is Boltzmann's constant. All the tests reported here use a temperature  $T = 300$  K. By choosing the origin and the orientation of the Cartesian coordinate system, we can set  $\vec{x}_{CM} = 0$  and  $\mathbf{R} = \mathbf{I}$  at  $t = 0$  without loss of generality. We choose  $\vec{v}_{CM}$  and  $\vec{\Omega}$  at  $t = 0$  to be vectors whose orientations are set with a random number generator and whose magnitudes are set by requiring the translational and rotational kinetic energies to satisfy

$$\frac{1}{2}kT = \frac{1}{2}M \vec{v}_{CM} \cdot \vec{v}_{CM}, \quad (42)$$

$$\frac{1}{2}kT = \frac{1}{2} \sum_A m_A \left[ \vec{x}_{0A} \cdot \vec{x}_{0A} \vec{\Omega} \cdot \vec{\Omega} - (\vec{x}_{0A} \cdot \vec{\Omega})^2 \right]. \quad (43)$$

The mode amplitudes  $\mathcal{A}_\mu$  and their time derivatives  $d\mathcal{A}_\mu/dt$  are chosen at  $t = 0$  to ensure that each normal mode of the molecule has energy  $kT$ ,

$$\mathcal{A}_\mu = \frac{1}{\omega_\mu} \sqrt{\frac{2kT}{M}} \sin \varphi_\mu, \quad (44)$$

$$\frac{d\mathcal{A}_\mu}{dt} = \sqrt{\frac{2kT}{M}} \cos \varphi_\mu, \quad (45)$$

where  $\varphi_\mu$  are randomly selected phases. These initial values for  $\{\vec{x}_{CM}, \vec{v}_{CM}, \mathbf{R}, \vec{\Omega}, \mathcal{A}_\mu, d\mathcal{A}_\mu/dt\}$  are used to start the evolutions of the exact mode-basis representation of the MD equations. We convert them to the equivalent Cartesian representation variables using Eqs. (14) and (15) and use those initial data to start our exact Cartesian MD evolutions. We also use the same initial values of  $\{\vec{x}_{CM}, \vec{v}_{CM}, \mathbf{R}, \vec{\Omega}, \mathcal{A}_\mu, d\mathcal{A}_\mu/dt\}$  to set the initial data for the SMA and MCSMA approximation tests. And finally we use these same values for  $\{\vec{x}_{CM}, \vec{v}_{CM}, \mathbf{R}, \vec{\Omega}\}$  with  $\mathcal{A}_\mu = d\mathcal{A}_\mu/dt = 0$  to set the initial data for the ZMA and MCZMA approximation tests.

All the MD simulation methods considered here consist of systems of ordinary differential equations of the form  $d\vec{Y}/dt = \vec{F}(\vec{Y}, t)$ , where  $\vec{Y}$  is the  $n$ -dimensional vector consisting of the dynamical fields in a particular method and  $\vec{F}(\vec{Y}, t)$  is the right side of the evolution equations for those fields. We solve these systems numerically with the initial data  $\vec{Y} = \vec{Y}(0)$  described above using an 8th order integrator by Dormand and Prince with dynamic time step size control (see the work of Hairer *et al.*<sup>27</sup> for details). This algorithm controls the error in  $\vec{Y}(t)$  by adjusting the time step size to keep an estimate of the local time-truncation error below  $\epsilon_\tau (|\vec{Y}| + 1)$  (see the work of Hairer *et al.*<sup>27</sup> for details about this time step control). We run each simulation with several values of the time step accuracy parameter  $\epsilon_\tau$  in the range  $10^{-13} \leq \epsilon_\tau \leq 10^{-6}$  to verify that time step errors are not the dominant cause of any differences we may see between the various MD evolution methods.

## B. Testing the exact mode-basis representation

Our first numerical tests of the model problem described in Sec. III A are designed to examine the differences between MD simulations performed with the standard Cartesian-basis representation of the MD equations, Eq. (1), and the exact mode-basis representation given in Eqs. (17)–(20). To perform these tests, we use the exact Cartesian-basis solution computed with time step accuracy parameter  $\epsilon_\tau = 10^{-13}$  as the reference solution with which to compare everything else. We refer to this reference solution as  $\vec{x}_A^{Ref}(t)$ .

We first measure how sensitively the exact Cartesian-basis MD evolutions depend on the time step accuracy parameter  $\epsilon_\tau$ . To do this, we evaluate the quantity  $\mathcal{E}_x$  that measures the differences between solutions computed using different time step accuracy parameters,  $\epsilon_\tau$ , and the reference solution,

$$\mathcal{E}_x(\epsilon_\tau) = \sqrt{\frac{1}{N} \sum_A |\vec{x}_A(\epsilon_\tau) - \vec{x}_A^{Ref}|^2}. \quad (46)$$

The solid curves in Fig. 2 show  $\mathcal{E}_x(\epsilon_\tau)$  for exact Cartesian-basis MD simulations of the  $C_{20}$  fullerene molecule computed with four different values of  $\epsilon_\tau = \{10^{-6}, 10^{-8}, 10^{-10}, 10^{-12}\}$ . Data points for these curves are obtained by evaluating  $\mathcal{E}_x$  at 10 ps time intervals during the evolutions. Each of these curves begins at early times with  $\mathcal{E}_x \approx \epsilon_\tau$  and then grows exponentially until  $\mathcal{E}_x \approx 1$  where it remains relatively constant for the duration of the simulation. These curves confirm the expectation that MD simulations are chaotic. Although the initial data at  $t = 0$  for these various runs are identical, after one



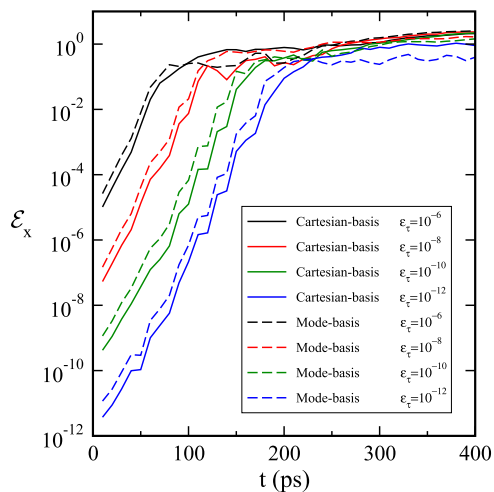


FIG. 2. Solid curves show  $\mathcal{E}_x$  computed with the exact Cartesian-basis MD code for different values of the time step accuracy parameter  $\epsilon_\tau$ . Dashed curves show  $\mathcal{E}_x$  for evolutions of the exact mode-basis MD code for different  $\epsilon_\tau$ . The exponential growth in these  $\mathcal{E}_x$  is caused by the chaotic nature of MD evolutions.

time step, the solutions differ from the one specified by the initial data, by amounts that depend on the time step accuracy parameter  $\epsilon_\tau$ . By definition, chaotic dynamical systems have the property that nearby solutions diverge exponentially. Figure 2 confirms that this is what is going on by showing that each of these evolutions of the fullerene  $C_{20}$  molecule diverges from the reference solution at the same exponential rate. The analogous graphs for the other fullerene molecules,  $C_{26}$ ,  $C_{60}$ , and  $C_{70}$ , included in our study are very similar, except for the time scale on which the chaotic instability occurs. In the  $C_{26}$  case, the instability grows at about half the rate of the  $C_{20}$  case, while the instabilities in the  $C_{60}$  and the  $C_{70}$  cases grow at two or three times the  $C_{20}$  rate. The presence of chaotic behavior in these MD simulations demonstrates why it is impossible to compute molecular evolutions in complete detail. Only certain macroscopic features of the evolutions, like dynamically conserved quantities such as the energy, momentum, and angular momentum are reproducible and simulatable.

The dashed curves in Fig. 2 show  $\mathcal{E}_x$  for simulations based on the exact mode-basis MD representation. We use the same Cartesian-basis reference solution  $\vec{x}_A^{Ref}$  when computing  $\mathcal{E}_x$  for these mode-basis simulations, and we see that the exponential divergence from the reference solution has exactly the same structure it has for the Cartesian-basis evolutions. The only difference is that the values of  $\mathcal{E}_x$  are somewhat larger,  $\mathcal{E}_x \approx 3\epsilon_\tau$ , at very early times in the mode-basis case. These differences appear to be caused by the fact that the mode-basis equations are much more complicated than their Cartesian-basis counterparts, so the truncation errors are somewhat higher in this case for fixed  $\epsilon_\tau$ . The rate of the exponential divergence from the reference solution is the same as the Cartesian-basis case, so Fig. 2 confirms that the mode-basis equations produce the same evolutions as the standard Cartesian-basis representation of MD. More importantly perhaps, these tests also confirm that our codes to evolve both versions of the MD equations contain no serious errors.

We have also monitored how well the energy  $E$  defined by

$$E = \frac{1}{2} \sum_A \frac{d\vec{x}_A}{dt} \cdot \frac{d\vec{x}_A}{dt} + U(\vec{x}_B), \quad (47)$$

the total momentum  $\vec{P}$  defined in Eq. (28), and the total angular momentum (about the center of mass)  $\vec{J}$  defined in Eq. (29) are conserved in these exact MD evolutions. In the absence of external forces (like van der Waals interactions with other molecules), these quantities should all be conserved by the exact MD evolution equations. To monitor these conserved quantities, we define  $\mathcal{E}_E$ ,  $\mathcal{E}_P$  and  $\mathcal{E}_J$  that measure the deviations of these quantities from their initial values

$$\mathcal{E}_E(t) = \frac{|E(t) - E(0)|}{E(0)}, \quad (48)$$

$$\mathcal{E}_P(t) = \frac{|\vec{P}(t) - \vec{P}(0)|}{|\vec{P}(0)|}, \quad (49)$$

$$\mathcal{E}_J(t) = \frac{|\vec{J}(t) - \vec{J}(0)|}{|\vec{J}(0)|}. \quad (50)$$

The solid curves in Fig. 3 show the evolution of these energy and momentum conservation errors for the simulations of the Cartesian-basis representation of the MD equations. The results in this figure represent those of the highest resolution simulations, i.e., those computed with time step error parameter  $\epsilon_\tau = 10^{-13}$ . We see from these figures that each of the conservation error quantities begins at small times with  $\mathcal{E}_E \approx \mathcal{E}_P \approx \mathcal{E}_J \approx \epsilon_\tau$ , which then grow slowly, roughly as a power law in time:  $\mathcal{E} \propto t^k$ , with  $k \lesssim 2$ . (The growth of truncation level errors in this way is typical of explicit numerical ordinary differential equation integrators such as the Dormand-Prince algorithm used in our tests.) The dashed curves in Fig. 3 show the errors in these conserved quantities for evolutions of the same initial data using the mode-basis representation of MD. The mode-basis results shown in Fig. 3 were also computed using the time step error parameter  $\epsilon_\tau = 10^{-13}$ . These results confirm that both versions of the

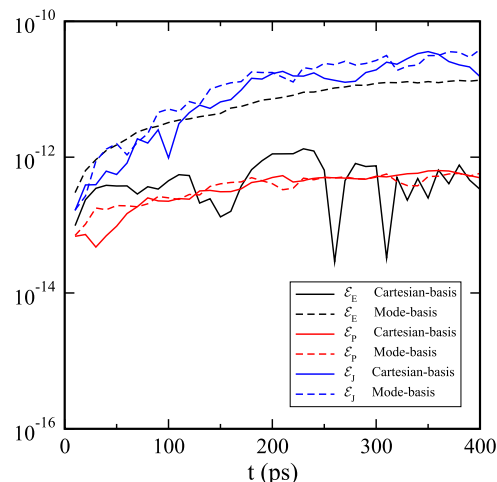


FIG. 3. Solid curves show  $\mathcal{E}_E$ ,  $\mathcal{E}_P$ , and  $\mathcal{E}_J$  for evolutions using the exact Cartesian MD code with time step accuracy  $\epsilon_\tau = 10^{-13}$ . Dashed curves show these same quantities for evolutions using the exact mode-basis MD code.

exact MD equations conserve the energy, the linear momentum, and the angular momentum of molecules at the level of the numerical truncation error used.

### C. Testing the large-scale approximations

In this section, we present the results of numerical tests of the large-scale MD approximations developed in Sec. II C. These approximations include the SMA and ZMA approximations that use a sinusoidal-in-time approximation (SMA) or the zero approximation (ZMA), respectively, for the mode amplitudes  $\mathcal{A}_\mu(t)$ . These approximations solve the exact mode-basis evolution Eqs. (17)–(19) for  $\{\vec{x}_{CM}, \vec{v}_{CM}, \mathbf{R}, \vec{\Omega}\}$  and simply ignore the exact evolution Eq. (20) for  $\mathcal{A}_\mu(t)$ . We also test approximations that use angular momentum conservation instead of Eq. (19) to determine the evolution of  $\vec{\Omega}(t)$ . The equations for these momentum conserving approximations, MCSMA and MCZMA, are given in Eqs. (37) and (38). All the numerical results shown here use the highest time resolution,  $\epsilon_\tau = 10^{-13}$ , in evolutions of the  $C_{20}$  fullerene molecule.

First we test how well these large-scale approximations conserve the energy  $E$ , linear momentum  $\vec{P}$ , and angular momentum  $\vec{J}$  during the evolutions of our model problem. We use the quantities  $\mathcal{E}_E$ ,  $\mathcal{E}_P$ , and  $\mathcal{E}_J$  defined in Eqs. (48)–(50) to monitor conservation violations. The solid curves in Fig. 4 show the energy conservation violations  $\mathcal{E}_E$  for the SMA and ZMA approximations, while the dashed curves show these violations for the MCSMA and MCZMA approximations. We see that the zero mode approximations ZMA and MCZMA conserve energy much better than the sinusoidal mode approximations SMA and MCSMA. However, even these sinusoidal mode approximations give energy conservation violations below the 0.1% level for these test problems.

The solid curves in Fig. 5 show the linear momentum conservation violations  $\mathcal{E}_P$  for the SMA and ZMA approximations, while the dashed curves show these violations for the MCSMA and MCZMA approximations. We see that linear momentum violations  $\mathcal{E}_P$  are much smaller for MCSMA than the SMA approximation, while its values are about the same

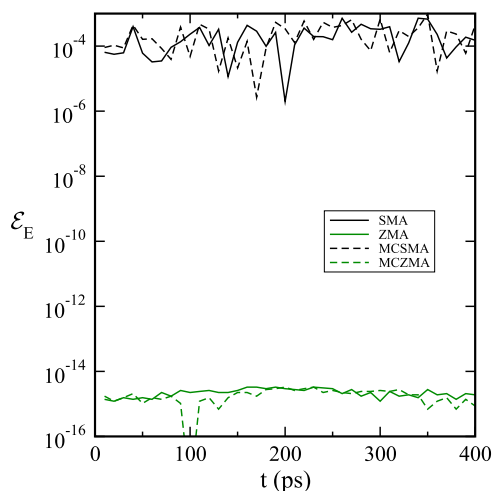


FIG. 4. Solid curves show energy conservation violations  $\mathcal{E}_E$  for the large-scale approximations SMA and ZMA, while dashed curves show these violations for the MCSMA and MCZMA approximations.

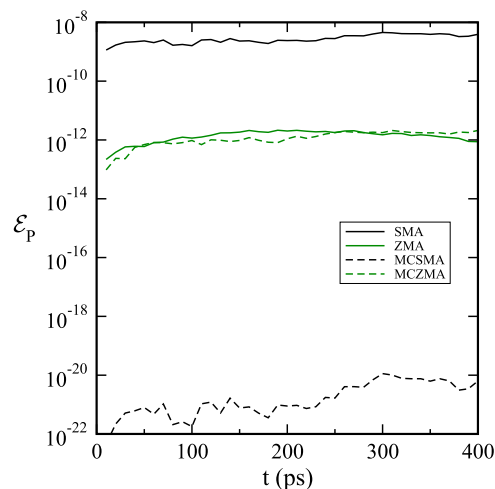


FIG. 5. Solid curves show linear momentum conservation violations  $\mathcal{E}_P$  for the large-scale approximations SMA and ZMA, while dashed curves show these violations for the MCSMA and MCZMA approximations.

for the MCZMA and ZMA approximations. Figure 5 shows, however, that linear momentum conservation is excellent for all of these large-scale approximations.

The solid curves in Fig. 6 show the angular momentum conservation violations  $\mathcal{E}_J$  for the SMA and ZMA approximations, while the dashed curves show these violations for the MCSMA and MCZMA approximations. We see that angular momentum conservation  $\mathcal{E}_J$  is much better for MCSMA than the SMA approximation, while these violations are about the same for the MCZMA and ZMA approximations. Not surprisingly, the momentum conserving approximation MCSMA has much better linear and angular momentum conservation properties than SMA, and somewhat better momentum conservation than the zero mode amplitude approximations MCZMA and ZMA.

Finally, we test how well the large-scale approximations reproduce the macroscopic variables  $\{\vec{x}_{CM}, \vec{v}_{CM}, \mathbf{R}, \vec{\Omega}\}$  for our model problem. We use the exact mode-basis representation with  $\epsilon_\tau = 10^{-13}$  as our reference solution in this

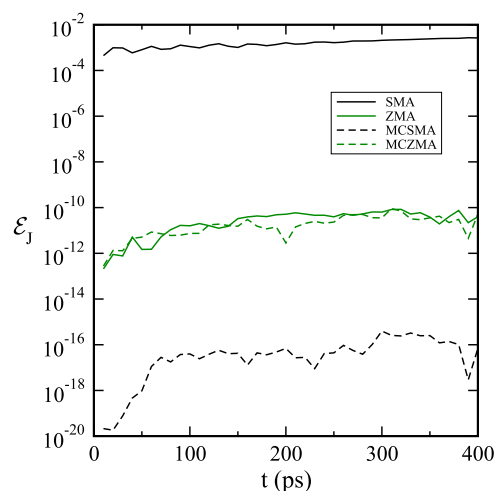


FIG. 6. Solid curves show angular momentum conservation violations  $\mathcal{E}_J$  for the large-scale approximations SMA and ZMA, while dashed curves show these violations for the MCSMA and MCZMA approximations.

case to test the various large-scale approximations. We evaluate the differences between the approximate and the exact solutions using the quantities  $\mathcal{E}_{x_{CM}}$ ,  $\mathcal{E}_{v_{CM}}$ ,  $\mathcal{E}_{\Omega}$ , and  $\mathcal{E}_q$ , defined by

$$\mathcal{E}_{x_{CM}} = |\vec{x}_{CM} - \vec{x}_{CM}^{Ref}|, \quad (51)$$

$$\mathcal{E}_{v_{CM}} = |\vec{v}_{CM} - \vec{v}_{CM}^{Ref}|, \quad (52)$$

$$\mathcal{E}_{\Omega} = \frac{|\vec{\Omega} - \vec{\Omega}^{Ref}|}{|\vec{\Omega}^{Ref}|}, \quad (53)$$

$$\mathcal{E}_q = \frac{1}{2} \sqrt{\sum_i (q_i - q_i^{Ref})^2}. \quad (54)$$

The solid curves in Fig. 7 show errors in the center of mass position,  $\mathcal{E}_{x_{CM}}$ , for the SMA and ZMA approximations, and the dashed curves show these errors for the MCSMA and MCZMA approximations. For comparison, the dotted curve in Fig. 7 shows the error in the somewhat lower resolution,  $\epsilon_{\tau} = 10^{-12}$ , evolution of the exact mode-basis representation compared to the reference solution. All of these errors are very small and are only growing slowly with time, approximately like  $\mathcal{E}_{x_{CM}} \propto t^2$ . These graphs confirm that all the large-scale approximations are able to determine  $\vec{x}_{CM}$  with excellent precision.

The solid curves in Fig. 8 show the errors in the velocity of the center of mass,  $\mathcal{E}_{v_{CM}}$ , for the SMA and ZMA approximations, and the dashed curves show these errors for the MCSMA and MCZMA approximations. For comparison, the dotted curve in Fig. 8 shows the error in the somewhat lower resolution,  $\epsilon_{\tau} = 10^{-12}$ , evolution of the exact mode-basis representation compared to the reference solution. All of these errors are very small and appear almost constant in time at late times. These graphs confirm that all the large-scale approximations are able to determine  $\vec{v}_{CM}$  with excellent precision.

The solid curves in Fig. 9 show the errors in the orientation matrix  $\mathbf{R}$ , as measured by  $\mathcal{E}_q$ , for the SMA and ZMA approximations, and the dashed curve shows these errors for

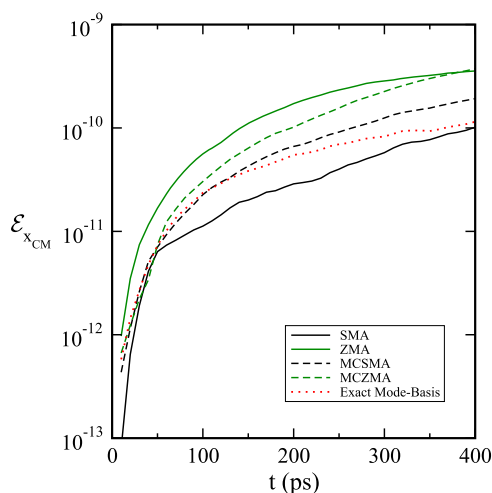


FIG. 7. Solid curves show errors in the center of mass position  $\mathcal{E}_{x_{CM}}$  for the large-scale approximations SMA and ZMA, while dashed curves show these errors for the MCSMA and MCZMA approximations. Dotted curve shows  $\mathcal{E}_{x_{CM}}$  for an exact mode-basis evolution with  $\epsilon_{\tau} = 10^{-12}$ .

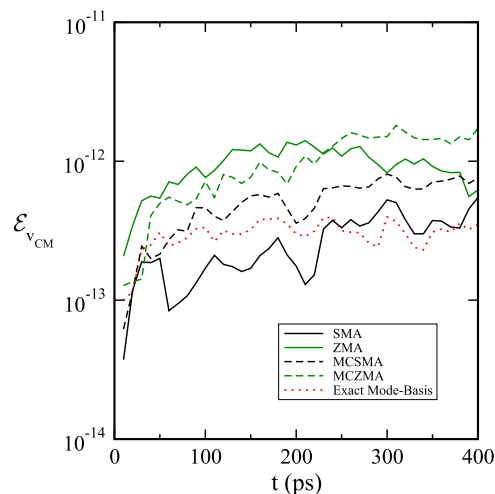


FIG. 8. Solid curves show errors in the center of mass velocity  $\mathcal{E}_{v_{CM}}$  for the large-scale approximations SMA and ZMA, while dashed curves show these errors for the MCSMA and MCZMA approximations. Dotted curve shows  $\mathcal{E}_{v_{CM}}$  for an exact mode-basis evolution with  $\epsilon_{\tau} = 10^{-12}$ .

the MCSMA approximation. The  $\mathcal{E}_q$  curve for the MCZMA approximation is indistinguishable from the ZMA curve. For comparison, the dotted curve in Fig. 9 shows the error in the somewhat lower resolution,  $\epsilon_{\tau} = 10^{-12}$ , evolution of the exact mode-basis representation compared to the reference solution. The exact dotted curve in Fig. 9 shows the exponential growth at early times which signals the presence of chaotic dynamics in this variable. The large scale approximations all have errors  $\mathcal{E}_q$  that are comparable to the late time behavior of the exact MD simulation. While it may be surprising that the orientation of the molecule cannot be predicted accurately from initial conditions of the molecule using an exact MD simulation, it is not surprising in this case that the large-scale approximations all give rather poor results for  $\mathbf{R}$  as well.

The solid curves in Fig. 10 show errors in  $\vec{\Omega}$ , as measured by  $\mathcal{E}_{\Omega}$ , for the SMA and ZMA approximations, and the

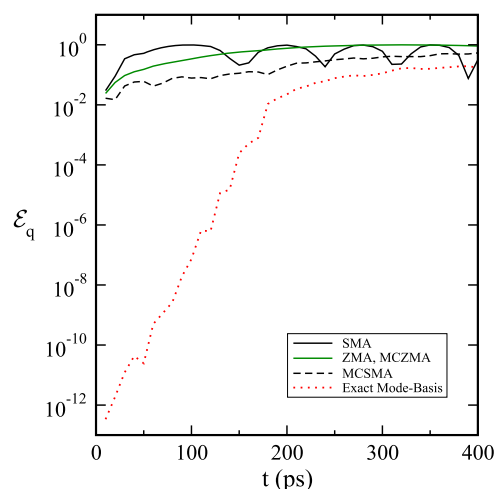


FIG. 9. Solid curves show errors in the orientation parameters  $\mathcal{E}_q$  for the large-scale approximations SMA and ZMA, while dashed curve shows these errors for the MCSMA approximation. The MCZMA errors are indistinguishable from the ZMA errors in this graph. Dotted curve shows  $\mathcal{E}_q$  for an exact mode-basis evolution with  $\epsilon_{\tau} = 10^{-12}$ .

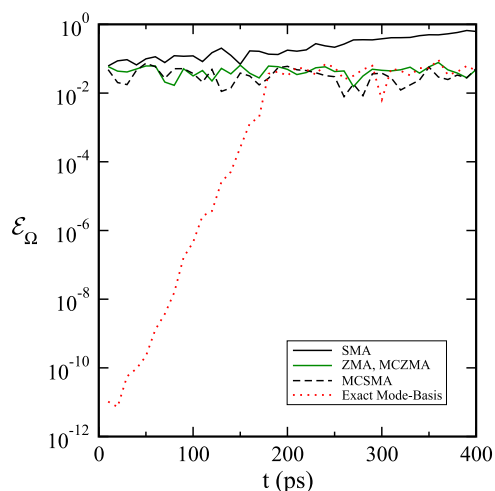


FIG. 10. Solid curves show errors in the angular velocity  $\mathcal{E}_\Omega$  for the large-scale approximations SMA and ZMA, while dashed curve shows these errors for the MCSMA approximation. The MCZMA errors are indistinguishable from the ZMA errors in this graph. Dotted curve shows  $\mathcal{E}_\Omega$  for an exact mode-basis evolution with  $\epsilon_\tau = 10^{-12}$ .

dashed curves show these errors for the MCSMA approximation. The MCZMA curve for  $\mathcal{E}_\Omega$  is indistinguishable from the ZMA curve. For comparison, the dotted curve in Fig. 10 shows the error in the somewhat lower resolution,  $\epsilon_\tau = 10^{-12}$ , evolution of the exact mode-basis representation compared to the reference solution. The exact dotted curve in Fig. 10 shows the exponential growth signaling the presence of chaotic dynamics in this variable. Given the chaos seen in the evolution of the orientation matrix  $\mathbf{R}$  seen in Fig. 9, it is not at all surprising that similar chaotic behavior is seen in the evolution of  $\vec{\Omega}$ . Thus it is not surprising that the large scale approximations all have errors  $\mathcal{E}_\Omega$  that are comparable to the late time behavior of the exact MD simulations.

In summary, all the large scale approximations do an excellent job of conserving linear momentum. All the large scale approximations except SMA do an excellent job of conserving angular momentum. The ZMA and MCZMA approximations do an excellent job of energy conservation, while the SMA and MCSMA approximations do not do so well. All of the large-scale approximations do excellent jobs of modeling  $\vec{x}_{CM}$  and  $\vec{v}_{CM}$ . None of the large scale approximations do a good job of modeling the macroscopic orientation variables  $\mathbf{R}$  and  $\vec{\Omega}$ . Overall then, our results show that the ZMA and the MCZMA approximations are more reliable approximations of the exact MD equations than the SMA and the MCSMA approximations.

#### D. Computational efficiency

This section briefly discusses the computational costs of the various MD evolutions used in our tests. Figure 11 shows the total run times (in seconds) for the exact Cartesian-basis simulation tests performed for each of the fullerene molecules  $C_{20}$ ,  $C_{26}$ ,  $C_{60}$ , and  $C_{70}$ . For each molecule, we ran five evolutions with  $\epsilon_\tau = \{10^{-6}, 10^{-8}, 10^{-10}, 10^{-12}, 10^{-13}\}$ , with each evolution simulating 400 ps of the molecular motion. Figure 11 shows that the run times for these tests increase

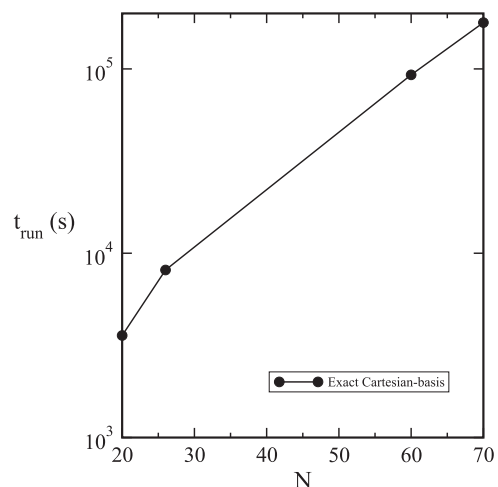


FIG. 11. Total run times in seconds for the 400 ps simulations of the fullerene molecules  $C_N$  using the Cartesian-basis version of the MD evolution code. Total run time includes the runs using five different values of the time step error parameters  $\epsilon_\tau = \{10^{-6}, 10^{-8}, 10^{-10}, 10^{-12}, 10^{-13}\}$ .

exponentially as the number of atoms in the simulation increases. A reasonably good approximation of these total run times is given by  $t_{\text{run}} \approx 1500 \times 10^{N/33}$ . The code we wrote to implement these methods was not highly optimized, so we expect that the computational efficiency could almost certainly be improved.

In Fig. 12, we illustrate the relative computational costs of performing evolutions using the various versions of the MD evolution equations discussed here. The solid curves connect the data points that represent the ratios of the total run times for the SMA and ZMA large-scale approximations to the total run time for the exact Cartesian-basis evolution. The dashed curves connect the analogous data points for the MCSMA and the MCZMA approximations. Finally, the dotted curve connects the data points that represent the ratios of the total

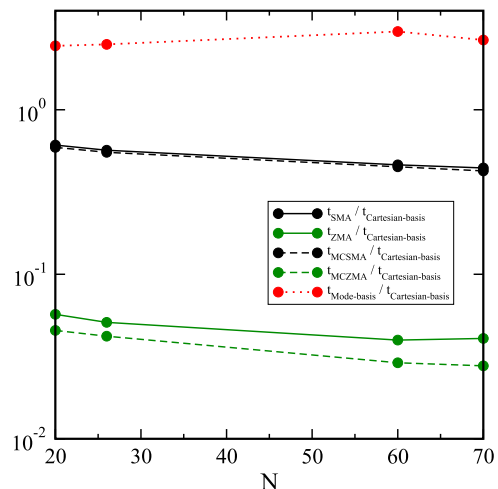


FIG. 12. Relative run times for the various large-scale MD approximations as functions of  $N$  the number of atoms in the simulation. Solid curves represent the ratios of the run times of the SMA and ZMA approximations with the run time for the exact Cartesian-basis representation. Dashed curves give the analogous ratios for the MCSMA and the MCZMA approximations. Dotted curve gives the ratio of the run times for the exact mode-basis representation to the exact Cartesian-basis representation.



run times for the exact mode-basis MD representation to the total run time for exact Cartesian-basis representation. These results show that computations using the SMA and MCSMA approximations are about twice as fast as those using the exact Cartesian-based representation, while the ZMA and MCZMA approximations are more than 25 times faster. The dotted curve in Fig. 12 shows that computations using the exact mode-basis representation of MD are two or three times slower than the exact Cartesian-basis representation. On the basis of computational efficiency, the MCZMA large-scale approximation is by far the best of the various MD simulation methods tested here.

#### IV. DISCUSSION

We have developed a new mode-based representation of the classical MD equations of motion that separate the macroscopic position and orientation degrees of freedom of a molecule from the internal vibrational degrees of freedom. We have confirmed through our numerical tests that most details of a molecular dynamical state evolve chaotically, including the large scale orientation and angular velocity of the molecule. Consequently those features cannot be predicted accurately even with exact MD simulations. We have derived a number of new large-scale approximations (based on our new mode-based representation) specifically designed to simulate accurately those features of MD evolutions that are not chaotic. We have shown through a series of careful numerical tests that some of these large-scale approximations give reliable, accurate predictions for the macroscopic properties of molecular motions, including the energies, linear and angular momentum, and the positions and velocities of their centers of mass. We find that one of these new approximations (MCZMA, the best of these new large-scale approximations studied here) is more than 25 times faster than our exact Cartesian-basis MD code, while giving comparable accuracy for the large-scale molecular properties. We also note that the MCZMA approximation does not depend on the mode basis vectors  $\vec{e}_A^\mu$  at all, which makes it very easy to implement numerically. Thus we conclude that there are many reasons to use reliable well-tested approximations for MD simulations rather than performing simulations using the full exact MD equations.

The tests done for this study focused on simulations of the dynamics of single molecules. The formalism created here has been designed, however, to accommodate simulations of collections of molecules in a straightforward way. The only change that needs to be made, as we noted earlier, is to change the single index  $A$  used to identify individual atoms to a pair of indices  $mA$ , where the first index  $m$  determines to which molecule the particular atom belongs. Macroscopic properties of the molecule like the energy,  $E$ , linear and angular momenta,  $\vec{P}$  and  $\vec{J}$ , and the position and velocity of the center of mass,  $\vec{x}_{CM}$  and  $\vec{v}_{CM}$ , should also acquire indices to identify to which molecule they belong:  $\{E_m, \vec{P}_m, \vec{J}_m, \vec{x}_{CMm}, \vec{v}_{CMm}\}$ . Then, by including van der Waals and/or Coulomb forces in the interaction potential  $U(\vec{x}_{mA})$ , it would be possible to study interactions between molecules using any of the large-scale approximations introduced here. The interactions modeled in

this way should be essentially identical to those interactions in an exact MD simulation.

#### ACKNOWLEDGMENTS

We thank Andrew McCammon and Lane Votapka for numerous helpful discussions on molecular dynamics. Funding and support for this work came from the National Biomedical Computation Resource through NIH Grant No. P41 GM103426.

#### APPENDIX A: NORMAL MODE BASIS

One natural choice for the mode-basis vectors  $\vec{e}_A^\mu$  is the eigenvectors of the Hessian of the potential energy function

$$0 = -m_A \omega_\mu^2 \vec{e}_A^\mu + \sum_B \frac{\partial^2 U}{\partial \vec{x}_B \partial \vec{x}_A} \cdot \vec{e}_B^\mu, \quad (\text{A1})$$

where  $\partial^2 U / \partial \vec{x}_B \partial \vec{x}_A$  in this equation is to be evaluated at the equilibrium state of the molecule where  $\vec{x}_A = \vec{x}_{0A}$ . Since the Hessian is a symmetric  $3N \times 3N$  dimensional real matrix, the eigenvalues  $\omega_\mu^2$  and eigenvectors  $\vec{e}_A^\mu$  are also real, and the collection of eigenvectors form a complete basis for the  $3N$  dimensional space of vectors. Equation (A1) is equivalent to Newton's equation of motion, Eq. (1), for the case of very small amplitude oscillations about its equilibrium state, so we call these  $\vec{e}_A^\mu$  the normal-mode basis. For stable molecules, the normal-mode frequencies are real, so all the eigenvalues of such systems are non-negative:  $\omega_\mu^2 \geq 0$ .

The zero-frequency modes of a molecule are represented by the eigenvectors of the Hessian matrix of the equilibrium potential energy function having zero eigenvalues,

$$0 = \sum_B \frac{\partial^2 U(\vec{x}_{0C})}{\partial \vec{x}_A \partial \vec{x}_B} \cdot \vec{e}_B^\mu. \quad (\text{A2})$$

We will now show that the eigenvectors corresponding to over-all rigid translations and rotations of the molecule are zero frequency modes.

We assume that the equilibrium state of the molecule is invariant under rigid spatial translations and rotations. This assumption makes sense to ensure that the equilibrium state of interest to us is the one where the molecule is isolated and does not interact with its large scale environment. If the molecule is invariant under translations, then the forces acting on the individual atoms must also be invariant. In its equilibrium state, the total force acting on each atom must vanish; therefore, the gradient of the potential energy function must vanish

$$0 = \frac{\partial U(\vec{x}_{0C})}{\partial \vec{x}_A} = \frac{\partial U(\vec{x}_{0C} + \lambda \vec{\tau})}{\partial \vec{x}_A}, \quad (\text{A3})$$

where  $\vec{\tau}$  is an arbitrary vector that describes the same translation for all the atoms in the molecule, and where  $\lambda$  is an arbitrary parameter that determines the magnitude of the translation. We can express the forces acting on each atom of a molecule that has been translated by an infinitesimal amount

using a Taylor series expansion

$$\begin{aligned} 0 &= \frac{\partial U(\vec{x}_{0C} + \lambda \vec{\tau})}{\partial \vec{x}_A} \\ &= \frac{\partial U(\vec{x}_{0C})}{\partial \vec{x}_A} + \lambda \sum_B \frac{\partial^2 U(\vec{x}_{0C})}{\partial \vec{x}_A \partial \vec{x}_B} \cdot \vec{\tau} + \mathcal{O}(\lambda^2). \end{aligned} \quad (\text{A4})$$

The first term on the right side of Eq. (A4) vanishes because of the equilibrium condition, Eq. (A3). Therefore, the second term on the right side of Eq. (A4) must also vanish for all values of  $\lambda$ . It follows that any vector  $\vec{\tau}$  that is the same for all the atoms in a molecule is a zero-frequency eigenvector,

$$0 = \sum_B \frac{\partial^2 U(\vec{x}_{0C})}{\partial \vec{x}_A \partial \vec{x}_B} \cdot \vec{\tau}. \quad (\text{A5})$$

The argument for the rotational invariance of the equilibrium state of the molecule is similar. Let  $\mathbf{R}(\lambda)$  denote a one-parameter family of rotation matrices. We assume that  $\lambda = 0$  corresponds to the identity rotation:  $\mathbf{R}(0) = \mathbf{I}$ . The rotational invariance of the equilibrium state of the molecule implies that

$$0 = \frac{\partial U[\vec{x}_{0C}]}{\partial \vec{x}_A} = \frac{\partial U[\mathbf{R}(\lambda) \cdot \vec{x}_{0C}]}{\partial \vec{x}_A}. \quad (\text{A6})$$

As before, we perform a Taylor expansion of the expression for the forces acting on an equilibrium molecule that has been rotated an infinitesimal amount,

$$0 = \frac{\partial U[\mathbf{R}(\lambda) \cdot \vec{x}_{0C}]}{\partial \vec{x}_A} \quad (\text{A7})$$

$$= \frac{\partial U[\vec{x}_{0C}]}{\partial \vec{x}_A} + \lambda \sum_B \frac{\partial^2 U[\vec{x}_{0C}]}{\partial \vec{x}_A \partial \vec{x}_B} \cdot \frac{d\mathbf{R}}{d\lambda} \Big|_{\lambda=0} \cdot \vec{x}_{0B} + \mathcal{O}(\lambda^2). \quad (\text{A8})$$

The derivative of any rotation matrix is an antisymmetric matrix. In this case, this matrix can be written as

$$\frac{dR_{ij}}{d\lambda} \Big|_{\lambda=0} = - \sum_k \epsilon_{ijk} \theta_k, \quad (\text{A9})$$

where the vector  $\vec{\theta}$  determines the direction and magnitude of the infinitesimal rotation. It follows that Eq. (A8) can be re-written as

$$0 = \frac{\partial U[\vec{x}_{0C}]}{\partial \vec{x}_A} + \lambda \sum_B \frac{\partial^2 U[\vec{x}_{0C}]}{\partial \vec{x}_A \partial \vec{x}_B} \cdot (\vec{\theta} \times \vec{x}_{0B}) + \mathcal{O}(\lambda^2). \quad (\text{A10})$$

The first term on the right side of Eq. (A10) vanishes because of the equilibrium condition, Eq. (A6). Therefore, the second term on the right side of Eq. (A10) must vanish for all values of  $\lambda$ . It follows that any vector of the form  $\vec{\theta} \times \vec{x}_{0A}$ , where  $\vec{\theta}$  is the same for all the atoms in a molecule, is a zero-frequency eigenvector

$$0 = \sum_B \frac{\partial^2 U[\vec{x}_{0C}]}{\partial \vec{x}_A \partial \vec{x}_B} \cdot (\vec{\theta} \times \vec{x}_{0B}). \quad (\text{A11})$$

The eigenvectors of the zero frequency modes (for generic molecules) therefore consist of rigid translations

$$\vec{e}_A^t(\vec{\tau}) = \vec{\tau}, \quad (\text{A12})$$

where  $\vec{\tau}$  is a constant vector that determines the magnitude and direction of the translation and rigid rotations,

$$\vec{e}_A^r(\vec{\theta}) = \vec{\theta} \times \vec{x}_{0A}, \quad (\text{A13})$$

where  $\vec{\theta}$  is a constant vector that determines the axis and magnitude of the rotation.

It is easy to show from Eq. (A1) that two eigenvectors,  $\vec{e}_A^\mu$  and  $\vec{e}_A^\nu$ , having different eigenvalues,  $\omega_\mu^2 \neq \omega_\nu^2$ , are orthogonal in the sense that,

$$0 = \sum_A \frac{m_A}{M} \vec{e}_A^\mu \cdot \vec{e}_A^\nu. \quad (\text{A14})$$

It follows that the translations  $\vec{e}_A^t(\vec{\tau}) = \vec{\tau}$  and rotations  $\vec{e}_A^r(\vec{\theta}) = \vec{\theta} \times \vec{x}_{0A}$  will be orthogonal to all of the non-zero frequency mode eigenvectors  $\vec{e}_A^\mu$ . These orthogonality conditions are given by

$$0 = \sum_A \frac{m_A}{M} \vec{e}_A^\mu \cdot \vec{e}_A^t(\vec{\tau}) = \sum_A \frac{m_A}{M} \vec{e}_A^\mu \cdot \vec{\tau}, \quad (\text{A15})$$

$$\begin{aligned} 0 &= \sum_A \frac{m_A}{M} \vec{e}_A^\mu \cdot \vec{e}_A^r(\vec{\theta}) \\ &= - \sum_A \frac{m_A}{M} (\vec{e}_A^\mu \times \vec{x}_{0A}) \cdot \vec{\theta}. \end{aligned} \quad (\text{A16})$$

These orthogonality conditions will hold for arbitrary values of the vectors  $\vec{\tau}$  and  $\vec{\theta}$ . Therefore these conditions can also be written in the form

$$0 = \sum_A \frac{m_A}{M} \vec{e}_A^\mu = \sum_A \frac{m_A}{M} \vec{e}_A^\mu \times \vec{x}_{0A}. \quad (\text{A17})$$

These conditions must hold for each non-zero frequency mode  $\mu$  and therefore demonstrate that the constraints on the mode-basis eigenvectors given in Eqs. (5) and (6) are satisfied by the normal-mode basis vectors.

## APPENDIX B: QUATERNION REPRESENTATION OF $\mathbf{R}(t)$

The differential equation that determines the rotation matrix  $\mathbf{R}(t)$ ,

$$\frac{dR_{ij}}{dt} = - \sum_{k\ell} \epsilon_{ik\ell} \Omega^\ell R_{kj}, \quad (\text{B1})$$

can be integrated numerically directly. Unfortunately the accumulation of truncation and roundoff errors in this direct approach inevitably produces a solution that is no longer a rotation matrix, and there is no reliable way to project out these errors to retrieve the correct  $\mathbf{R}(t)$ . A better approach is to adopt some parametric representation of the three-dimensional space of rotation matrices and then to convert Eq. (B1) into a system of equations for the evolution of those parameters, and finally to integrate that parametric representation numerically. For example, one common representation of the rotation matrices uses the Euler angles as parameters. Since the Euler angle representation is not one to one (at a few singular points), a better representation uses unit quaternions which do provide a one to one representation. We use the quaternion representation for our numerical solutions of  $\mathbf{R}(t)$ . Let  $q_0$  represent the real part and  $q_1$ ,  $q_2$ , and  $q_3$  represent the three independent imaginary parts of a quaternion with

$$q_0^2(t) + q_1^2(t) + q_2^2(t) + q_3^2(t) = 1. \quad (\text{B2})$$

This equation defines a unit three-sphere in this parameter space, so the space of possible parameter values is three-dimensional. A general rotation matrix  $\mathbf{R}$  can be written in terms of these quaternion parameters in the following way:

$$\mathbf{R} = 2 \begin{pmatrix} q_0^2 + q_1^2 - \frac{1}{2} & q_1 q_2 - q_0 q_3 & q_1 q_3 + q_0 q_2 \\ q_1 q_2 + q_0 q_3 & q_0^2 + q_2^2 - \frac{1}{2} & q_2 q_3 - q_0 q_1 \\ q_1 q_3 - q_0 q_2 & q_2 q_3 + q_0 q_1 & q_0^2 + q_3^2 - \frac{1}{2} \end{pmatrix}. \quad (\text{B3})$$

It is then straightforward to transform the rotation matrix evolution Eq. (B1) into an equation for the evolution of the quaternion parameters. The result is

$$\frac{dq_0(t)}{dt} = -\frac{1}{2}(\Omega_x q_1 + \Omega_y q_2 + \Omega_z q_3), \quad (\text{B4})$$

$$\frac{dq_1(t)}{dt} = \frac{1}{2}(\Omega_x q_0 + \Omega_y q_3 - \Omega_z q_2), \quad (\text{B5})$$

$$\frac{dq_2(t)}{dt} = \frac{1}{2}(-\Omega_x q_3 + \Omega_y q_0 + \Omega_z q_1), \quad (\text{B6})$$

$$\frac{dq_3(t)}{dt} = \frac{1}{2}(\Omega_x q_2 - \Omega_y q_1 + \Omega_z q_0). \quad (\text{B7})$$

Given a solution to these equations for  $q_0(t)$ ,  $q_1(t)$ ,  $q_2(t)$ , and  $q_3(t)$ , it is easy to reconstruct the rotation matrix  $\mathbf{R}(t)$  using Eq. (B3).

The constraint,

$$\mathcal{C} \equiv q_0^2 + q_1^2 + q_2^2 + q_3^2 - 1, \quad (\text{B8})$$

which measures how well the quaternion parameters remain on the unit three-sphere, is preserved by the evolution defined by Eqs. (B4)–(B7). In particular, these evolution equations imply

$$\frac{d\mathcal{C}}{dt} = 0. \quad (\text{B9})$$

Solving Eqs. (B4)–(B7) numerically will nevertheless result in truncation level violations of this constraint, so it is necessary to re-scale the solution periodically to ensure that  $\mathcal{C} = 0$ . Without this re-scaling,  $\mathbf{R}(t)$  constructed using Eq. (B3) will not be a rotation matrix.

The numerical solution of Eqs. (B4)–(B7) can be improved by adding constraint damping terms to the system. These extra terms vanish whenever the constraints are satisfied, thus leaving the desired solutions unchanged. But these constraint damping terms are chosen to drive the solution back toward the constraint satisfying surface whenever small numerical constraint violations inevitably occur. The resulting quaternion evolution equation with constraint damping is given by

$$\frac{dq_0(t)}{dt} = -\frac{1}{2}(\Omega_x q_1 + \Omega_y q_2 + \Omega_z q_3) - \frac{1}{8} \eta q_0 \mathcal{C}, \quad (\text{B10})$$

$$\frac{dq_1(t)}{dt} = \frac{1}{2}(\Omega_x q_0 + \Omega_y q_3 - \Omega_z q_2) - \frac{1}{8} \eta q_1 \mathcal{C}, \quad (\text{B11})$$

$$\frac{dq_2(t)}{dt} = \frac{1}{2}(-\Omega_x q_3 + \Omega_y q_0 + \Omega_z q_1) - \frac{1}{8} \eta q_2 \mathcal{C}, \quad (\text{B12})$$

$$\frac{dq_3(t)}{dt} = \frac{1}{2}(\Omega_x q_2 - \Omega_y q_1 + \Omega_z q_0) - \frac{1}{8} \eta q_3 \mathcal{C}. \quad (\text{B13})$$

These equations imply the following evolution equation for the constraint:

$$\frac{d\mathcal{C}}{dt} = -\eta(\mathcal{C} + 1)\mathcal{C}, \quad (\text{B14})$$

which drives constraint violations  $\mathcal{C}$  toward zero exponentially on a time scale set by the constant  $\eta$  when  $\eta > 0$ . This is the form of the rotation matrix evolution equations used in all of our numerical solutions of the various representations of the MD equations.

<sup>1</sup>J. A. McCammon and S. C. Harvey, *Dynamics of Proteins and Nucleic Acids* (Cambridge University Press, 1987).

<sup>2</sup>M. Allen and D. Tildesley, *Computer Simulation of Liquids* (Oxford Science Publications, 1987).

<sup>3</sup>D. Frenkel and B. Smit, *Understanding Molecular Simulation*, 2nd ed. (Academic Press, 2002).

<sup>4</sup>M. Karplus and J. A. McCammon, *Nat. Struct. Biol.* **9**, 646 (2002).

<sup>5</sup>R. A. Laskowski, F. Gerick, and J. M. Thornton, *FEBS Lett.* **583**, 1692 (2009), Prague Special Issue: Functional Genomics and Proteomics.

<sup>6</sup>J. D. Durrant and J. A. McCammon, *BMC Biol.* **9**, 1 (2011).

<sup>7</sup>X. Li, *Applications of Molecular Dynamics in Atmospheric and Solution Chemistry* (KTH Royal Institute of Technology, Stockholm, 2011).

<sup>8</sup>J. Wang and T. Hou, *J. Chem. Theory Comput.* **7**, 2151 (2011).

<sup>9</sup>M. O. Steinhauser, "Introduction to molecular dynamics simulations: Applications in hard and soft condensed matter physics," in *Molecular Dynamics—Studies of Synthetic and Biological Macromolecules* (InTech, 2012), pp. 3–28.

<sup>10</sup>J. R. Perilla, B. C. Goh, C. K. Cassidy, B. Liu, R. C. Bernardi, T. Rudack, H. Yu, Z. Wu, and K. Schulten, *Curr. Opin. Struct. Biol.* **31**, 64 (2015).

<sup>11</sup>G. Vettoretti, E. Moroni, S. Sattin, J. Tao, D. Agard, A. Bernardi, and G. Colombo, *Sci. Rep.* **6**, 23830 (2016).

<sup>12</sup>H. Poincaré, *Acta Math.* **13**, 1270 (1890).

<sup>13</sup>S. D. Bond and B. J. Leimkuhler, *Acta Numer.* **16**, 1 (2007).

<sup>14</sup>B. Leimkuhler and S. Reich, *Simulating Hamiltonian Dynamics* (Cambridge University Press, 2004).

<sup>15</sup>E. B. Wilson, Jr., *J. Chem. Phys.* **9**, 76 (1941).

<sup>16</sup>J.-P. Ryckaert, G. Ciccotti, and H. J. Berendsen, *J. Comput. Phys.* **23**, 327 (1977).

<sup>17</sup>H. C. Andersen, *J. Comput. Phys.* **52**, 24 (1983).

<sup>18</sup>R. C. Smith, *Uncertainty Quantification* (SIAM, 2014).

<sup>19</sup>P. N. Patrone, A. Dienstfrey, A. R. Browning, S. Tucker, and S. Christensen, *Polymer* **87**, 246 (2016).

<sup>20</sup>S. T. Reeve and A. Strachan, *J. Comput. Phys.* **334**, 207 (2017).

<sup>21</sup>F. Rizzi, H. N. Najm, B. J. Debuschere, K. Sargsyan, M. Salloum, H. Adalsteinsson, and O. M. Knio, *Multiscale Model. Simul.* **10**, 1428 (2012).

<sup>22</sup>F. Rizzi, H. N. Najm, B. J. Debuschere, K. Sargsyan, M. Salloum, H. Adalsteinsson, and O. M. Knio, *Multiscale Model. Simul.* **10**, 1460 (2012).

<sup>23</sup>F. C. Grogan, "Comp. tech. In MD and Det. Shock Dynamics," Ph.D. thesis, University of California at San Diego, 2017.

<sup>24</sup>M. Levitt, C. Sander, and P. S. Stern, *J. Mol. Biol.* **181**, 423 (1985).

<sup>25</sup>A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus, *J. Phys. Chem. B* **102**, 3586 (1998).

<sup>26</sup>W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in FORTRAN*, 2nd ed. (Cambridge University Press, Cambridge, England, 1992).

<sup>27</sup>E. Hairer, S. P. Nørsett, and G. Wanner, *Solving Ordinary Differential Equations I* (Springer, 2008).

<sup>28</sup>We note that for MD simulations of collections of molecules, the single index  $A$  that identifies the atoms should be replaced by a pair of indices  $mA$ , with  $m$  identifying the particular molecule and  $A$  the atom within that molecule. All the macroscopic properties of these molecules should also acquire an additional  $m$  index to identify which molecule the attribute belongs to, e.g.,  $M$  should become  $M_m$ ,  $\vec{x}_{CM}(t)$  should become  $\vec{x}_{CMm}(t)$ , etc. For simplicity of notation, we will suppress these molecule-identifying indices in the discussion in this paper.

<sup>29</sup>The angular momentum  $\vec{J}$  defined in this way is the angular momentum relative to the center of mass of the molecule.