Implicit time integration for multiscale molecular dynamics using transcendental Padé approximants

Andrew Abi Mansour, and Peter J Ortoleva

*J. Chem. Theory Comput.*, **Just Accepted Manuscript** • DOI: 10.1021/acs.jctc.5b01232 • Publication Date (Web): 04 Feb 2016

Downloaded from [http://pubs.acs.org](http://pubs.acs.org) on March 2, 2016

**Just Accepted**

“Just Accepted” manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides “Just Accepted” as a free service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. “Just Accepted” manuscripts appear in full in PDF format accompanied by an HTML abstract. “Just Accepted” manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are accessible to all readers and citable by the Digital Object Identifier (DOI®). “Just Accepted” is an optional service offered to authors. Therefore, the “Just Accepted” Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the “Just Accepted” Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these “Just Accepted” manuscripts.
Implicit time integration for multiscale molecular dynamics using transcendental Padé approximants

Andrew Abi Mansour and Peter J. Ortoleva*

Department of Chemistry and Center for Theoretical and Computational Nanoscience, Indiana University, Bloomington, IN 47405

E-mail: ortoleva@indiana.edu

Abstract

Molecular dynamics systems evolve through the interplay of collective and localized disturbances. As a practical consequence, there is a restriction on the timestep imposed by the broad spectrum of timescales involved. To resolve this restriction, multiscale factorization was introduced for molecular dynamics as a method that exploits the separation of timescales by coevolving the coarse-grained and atom-resolved states via Trotter factorization. Developing a stable time-marching scheme for this coevolution, however, is challenging because the coarse-grained dynamical equations depend on the microstate; therefore, these equations cannot be expressed in closed form. The objective of this paper is to develop an implicit time integration scheme for multiscale simulation of large systems over long periods of time and with high accuracy. The scheme uses Padé approximants to account for both the stochastic and deterministic features of the coarse-grained dynamics. The method is demonstrated for a protein...
either undergoing a conformational change or migrating under the influence of an ex-
ternal force. The method shows promise in accelerating multiscale molecular dynamics 
without loss of atomic precision or the need to conjecture the form of coarse-grained 
governing equations.

Introduction

Molecular dynamics (MD) is a computational method that has been widely employed to 
study the dynamics and structure of macromolecules at the atomic scale.\(^1\) However, simu-
lating supramolecular assemblies and other large nanosystems becomes a challenge for MD 
because of the broad spectrum of timescales involved. The objective of this study is to 
introduce a methodology that addresses this challenge. The method is based on multiscale 
factorization (MF),\(^2\)\(^-\)\(^4\) an equation-free method\(^5\)\(^,\)\(^6\) that uses Lie-Trotter factorization\(^7\) to co-
evolve atomistic and coarse-grained (CG) variables in time. In MF, the CG dynamics are 
governed by differential equations that are coupled to the atomistic dynamics and therefore 
cannot be expressed in closed form.

There are a variety of methods for solving time-dependent differential equations. Explicit 
methods calculate the state of a system at a later time from the state of the system at the 
current and previous times. Explicit methods are easy to implement because they often 
lead to discretized linear algebraic equations. However, they are not efficient for solving 
stiff differential equations.\(^8\) In contrast, implicit methods calculate the state of the system 
from both the current (or previous) state of the system and the future one. Although this 
usually involves solving non-linear equations, implicit methods are efficient in integrating 
stiff differential equations because they enable much larger timesteps to be taken.\(^8\)\(^-\)\(^10\)

In earlier studies,\(^2\)\(^,\)\(^3\)\(^,\)\(^11\)\(^,\)\(^12\) an explicit scheme (projective forward Euler (PFE))\(^13\) was used 
in MF to integrate in time the ensuing CG dynamical equations. However, PFE is is a first 
order method with limited accuracy and conditional stability. Omnipresent stochasticity 
(from the underlying atomistic dynamics) further restricts the CG timestep that can be taken
with PFE. History-based explicit time integration methods\textsuperscript{13,14} with extended stability region relative to PFE were introduced to address the separation of timescales in complex systems; these methods enable larger CG timesteps to be taken without compromising accuracy.

In the present study, an implicit scheme based on Padé approximants (PAs) is developed for solving stochastic differential equations that arise in MF and other equation-free methods. PAs provide a way to integrating known limiting mathematical behaviors into the solution of a problem, and notably expanding the range of applicability of an approximation.\textsuperscript{15} A well known example in statistical mechanics is the use of PAs in low density expansions of an $N$–particle system to predict gas to liquid phase transitions.\textsuperscript{16,17} Other examples are found in electrostatics for developing solutions of the Poisson-Boltzman equation\textsuperscript{18} or constructing solutions to the reaction diffusion problem in the context of nonlinear wave propagation.\textsuperscript{19,20}

The form of the PA used in this work is inspired by Ito’s theory;\textsuperscript{21} in this way, the present formulation accounts for both the coherent and stochastic influences on CG evolution. While the algorithm is explored in the context of multiscale molecular dynamics, it is applicable to a wider class of problems described by differential equations even when the closed form of which is not known. In addition to seeking an efficient atom-resolved multiscale MD algorithm, the objective of this study is to arrive at a framework that naturally separates the stochastic from the coherent dynamics of the problem. Therefore, the algorithm can be used as a separation technique for extracting the coherent facets of the dynamics of the problem.

The method introduced in this paper is referred to hereafter as multiscale factorization implicit (MFI). The theoretical development of which is discussed in section II. MFI is demonstrated in section III for a protein undergoing a conformational change or migrating under the influence of an externally forced flow in the host fluid. Finally, conclusions are drawn in section IV.
Formalism

Consider a macromolecular system comprising $N$ atoms and slowly evolving in time, such as a protein undergoing a conformational change or a virus-like particle expanding or shrinking in response to an external stimulus. The system is represented by $\Gamma$, the set of positions and momenta of all $N$ atoms corresponding to a specific microstate. In principle, $\Gamma$ contains all the relevant information for studying the dynamics of the problem. However, MD becomes inefficient for simulating large systems characterized by long timescales. A set of variables, $\phi$, is introduced to reduce the dimensionality of the problem and hence provide an efficient way for simulating these systems. While $\Gamma$ captures the atomistic fluctuations, $\phi$ captures the coherent and slowly varying changes, such as structural transitions or self-assembly, macromolecular systems undergo. In MD, the dynamics of $\Gamma$ are governed by Newton’s 2nd law, while $\phi$ follows a differential equation that results from multiscale factorization (MF)\(^2\) and takes the form

$$\frac{d\phi}{dt} = \Pi(\Gamma),$$  

(1)

where $\Pi$ is the CG velocity and $t$ represents time. In the MF method, $\Pi(\Gamma)$ was computed from the atomistic microphase. It was further shown\(^2,3\) that for many molecular systems, $\Pi$ constitutes a stationary process such that over a small period of time, $\delta$, much smaller than the characteristic time of CG dynamics, $\Delta$, the integral $\frac{1}{\delta} \int_t^{t+\delta} \Pi(t')dt'$ rapidly converges to $\frac{1}{\Delta} \int_t^{t+\Delta} \Pi(t')dt'$. Thus, the microstate represented by $\Gamma$ and the CG state represented by $\phi$ are coevolved in time such that a short MD run is initiated (for a period $\delta$), followed by the CG advancement in time over a period $\Delta$. The speedup obtained with this approach is therefore $\Delta/\delta$. Earlier,\(^2,3\) the CG time advancement was achieved via the projective forward Euler (PFE) method such that

$$\phi(t + \Delta) = \phi(t) + \frac{\Delta}{\delta}(\phi(t + \delta) - \phi(t)).$$  

(2)
This multiscale formalism was cast as MF using PFE and is discussed elsewhere in more detail.\textsuperscript{2,3}

PFE restricts the CG timestep $\Delta$ because of the former’s limited stability and accuracy. An implicit integration scheme, that provides improved efficiency and stability over PFE, is here introduced based on Padé approximants (PAs). The PA used here is

$$\phi(t - \Delta + \tau)^{PA} = \frac{a_0 + a_{1/2} \tau^{1/2} + a_1 \tau + a_{3/2} \tau^{3/2} + O(\tau^m)}{\sum_{i=0}^{n} b_i \tau^i},$$  \hspace{1cm} (3)$$

where $\tau$ is time relative to $t - \Delta$ such that $\tau \in [0, 2\Delta]$. This relative time framework facilitates the use of historical and future information in a symmetric way. The $a$ and $b$ Padé coefficients in Eq. (3) depend on the central time $t$ through their calibration using data before, at, and beyond $t$. The PA in Eq. (3) must have two essential properties:

- For a finite $\Delta$, $\phi^{PA} \left[ \frac{m}{n} \right]$ should be well-behaved for $\tau \in [0, 2\Delta]$,

- $\phi^{PA} \left[ \frac{m}{n} \right]$ should be consistent with the Ito formula\textsuperscript{21} which accounts for the noise from the underlying atomistic dynamics.

The PA in Eq. (3) could diverge unless there’s a restriction on the $b_i$ coefficients. Replacing the series in the denominator with an exponential term is one plausible way that avoids this divergence. Therefore, the simplest PA we found to satisfy the two aforementioned properties is

$$\phi(t - \Delta + \tau)^{PA} = \frac{a_0 + a_{1/2} \tau^{1/2} + a_1 \tau}{e^{b_1 \tau}}.$$

Equation (4) can be thought of as a transcendental Padé approximant (TPA) that reduces to the Euler-Maruyama formula\textsuperscript{22} when $b_1 = 0$. Expanding the denominator in terms of $\tau$ yields

$$\phi(t - \Delta + \tau)^{PA} = a_0 + a_{1/2} \tau^{1/2} + (a_1 - a_0 b_1) \tau + O(\tau^{3/2}).$$

Comparing Eq. (5) with the Euler–Maruyama formula, the deterministic part is $(a_1 - a_0 b_1) \tau$, while the noise is quantified by the $a_{1/2} \tau^{1/2}$ term. This facilitates an accurate and stable
computation of the CG velocity, \( \Pi \), which is needed for inertial multiscale problems. The TPA formula introduced here yields two types of time integration schemes discussed below.

### Explicit scheme

This method has a limited stability region, and it involves solving a set of four equations:

\[
\phi(t - \Delta) = a_0, \quad (6)
\]

\[
\phi(t - \Delta + \delta)e^{b_1\delta} = a_0 + a_{1/2}\delta^{1/2} + a_1\delta, \quad (7)
\]

\[
\phi(t)e^{b_1\Delta} = a_0 + a_{1/2}\Delta^{1/2} + a_1\Delta, \quad (8)
\]

\[
\phi(t + \delta)e^{b_1(\Delta + \delta)} = a_0 + a_{1/2}(\Delta + \delta)^{1/2} + a_1(\Delta + \delta). \quad (9)
\]

The Padé coefficients \((a_0, a_{1/2}, a_1, b_1)\) are the unknown variables, and they depend only on the past, i.e. on \(\phi(t - \Delta)\), \(\phi(t - \Delta + \delta)\), and \(\phi(t + \delta)\) (Fig. (1)). The unknown \(\phi(t + \Delta)\) is computed via Eq. (4) once the Padé coefficients are determined. This method serves as a predictor for estimating the Padé coefficients used in the implicit scheme discussed below.

### Implicit scheme

In this case, the Padé coefficients \((a_0, a_{1/2}, a_1, b_1)\) depend on both the past and future CG states, i.e. they are a function of \(\phi(t + \Delta)\) as well. In addition to the four equations included in the explicit scheme, a 5\textsuperscript{th} (self-consistency) equation that includes the future CG state,
must be simultaneously solved as well, i.e.

\[ \phi(t - \Delta) = a_0, \]  
\[ \phi(t - \Delta + \delta)e^{b_1 \delta} = a_0 + a_{1/2} \delta^{1/2} + a_1 \delta, \]  
\[ \phi(t)e^{b_1 \Delta} = a_0 + a_{1/2} \Delta^{1/2} + a_1 \Delta, \]  
\[ \phi(t + \delta)e^{b_1 (\Delta + \delta)} = a_0 + a_{1/2} (\Delta + \delta)^{1/2} + a_1 (\Delta + \delta), \]  
\[ \phi(t + \Delta)e^{b_1 (2\Delta)} = a_0 + a_{1/2} (2\Delta)^{1/2} + a_1 (2\Delta). \]  

(10)  
(11)  
(12)  
(13)  
(14)

The additional computations required for the implicit scheme yield an extended stability region as demonstrated in the next section.

Demonstration

The objective of this section is to assess the feasibility of MFI, which uses the TPA integration scheme discussed in the previous section, relative to MF which is based on PFE, as a viable multiscale MD algorithm. The MF scheme is characterized by two phases: the all-atom microphase of length \( \delta \), and the CG phase of length \( \Delta \). Here MFI is demonstrated for the CG phase of the computation.

Methodology

There are two sources of error that specifically follow from MF-MFI: 1) fine-graining error due to microstate reconstruction,\(^{11,23}\) and 2) coarse-grained error due to advancing the CG state in time. To assess the efficiency and accuracy of the latter in MFI, two examples are used. From each simulation, a time series for \( \phi(n\delta) \) was generated where \( n \) labels a discrete microphase (MD) run. This series was used to compute \( \phi(n\Delta) \) at selected CG timesteps. With this validation approach, the microphase is uncorrelated from the CG phase. This provides an accurate assessment of the time integration methods employed without any
finegraining error. Thus, $\phi(n\delta)$ is unaffected by the extrapolated values of $\phi(n\Delta)$. However, the propagation of error that results from the CG time integration of $\phi(n\Delta)$ is accounted for, and it is analyzed below. For MFI, a predictor-corrector method is used to compute $\phi(t + \Delta)$. In the predictor stage, Eqs. (6 - 9) of the explicit scheme were solved using Newton’s method to estimate the Padé coefficients, with all coefficients taken to be zero as an initial guess. In the corrector stage, Eqs. (10 - 14) were solved using Newton’s method, with the Padé coefficients and $\phi(t + \Delta)$ predicted in the first stage used as an initial guess.

Results and Discussion

In this subsection, we show how the MFI scheme provides enhanced stability over MF and speedup over MD. Density field variables at selected nodal points are used as CG variables. In particular, $\phi_\lambda$ for node $l$ discretized in space at $s_l$ is defined to be the field variable

$$\phi_\lambda(s_l) = \sum_{i=1}^{N} \lambda_i K(r_i - s_l, w),$$  \hspace{1cm} (15)

where $r_i$ is a 3D position vector of atom $i$, $\lambda_i$ a property of atom $i$ (such as its mass $m_i$, momentum $p_i$, etc.), $K(r_i - s_l, w)$ is the kernel whose width, $w$, impacts the resolution of the CG description, and $s_l$ is a 3D positional vector for node $l$. Specifically, the kernel is taken to be a normalized Gaussian function

$$K(r_i - s_l, w) = \frac{1}{w^3 \pi^{3/2}} e^{-\frac{(r_i - s_l)^T(r_i - s_l)}{w^2}},$$  \hspace{1cm} (16)

such that the summation in Eq. (15) is truncated when the kernel is below 0.01Å. Eq. (15) has been widely employed in smoothed particle hydrodynamics to simulate fluid flow at the macroscopic scale.\textsuperscript{24}

MFI is demonstrated for two different systems as discussed in the remainder part of this subsection. The microphase duration $\delta$ was set to 10 fs for the mass density field variable $\phi_m$ used to capture protein conformational change, and $\delta$ was set to 1 ps for the
momentum density field variable $\phi_p$ used to capture protein migration. Various values of the CG timestep $\Delta$ were used to assess the stability and accuracy for MF (using PFE) and MFI (using TPA). All simulations were performed with Gromacs$^{25-27}$ (v 5.1.0) under NVT conditions in aqueous solution using the CHARMM27 force field$^{28}$ and the TIP3P water model$^{29}$ to account for all intermolecular forces. Periodic boundary conditions were imposed on the system, and a 4$^{th}$ order particle mesh Ewald method was used to compute Coulomb forces.

**Protein structural transition**

Pertussis toxin (PDB ID: 1PRT), a protein-based exotoxin$^{30}$ (Fig. (2)), was used as an illustrative example. This protein was simulated at 300 K. NaCl counter-ions of concentration 0.15 M were added for charge neutrality. The system consisted of 603,775 atoms in a box of dimensions $[0, 16]$ nm, $[0, 16]$ nm, and $[0, 24]$ nm along the $x$, $y$, and $z$ axes, respectively. An equilibration run with position restraints imposed on the protein was performed for 100 ps; after thermal equilibrium was established, the system was simulated without any restraints for 2 ns during which the protein underwent a conformational change involving local increases and decreases in mass density (i.e. compression-extension).

A grid comprising $20 \times 20 \times 20$ nodal points was used to capture the spatial distribution of protein compaction and expansion. This yielded a total of 8000 field variables, of which a subset was used for analysis. In particular, the mass density $\phi_m$ at $s_l = [2.75, 4.76, 7.80]$ nm was used for analysis. This CG variable evolves on the picosecond timescale and is defined to be

$$\phi_m(r, s_l) = \sum_{i=1}^{N} m_i K(r_i - s_l, w),$$

where the width $w$, was set to 15 Å. The width chosen for $\phi_m$ ensured that it is a stochastic variable that slowly changes in time but contains a degree of fluctuations from the underlying atom-resolved dynamics. MF and MFI were used to reproduce the temporal profile of $\phi_m$, which captures mass transfer associated with protein deformation in Fig. (4). For $\Delta = 0.2$
ps, MF becomes unstable because the truncation error in PFE increases in time. This is due to the limited stability region of the PFE method. MF, on the other hand, accurately reproduces the trajectory of $\phi_m$ for $\Delta = 0.2$ ps and $\Delta = 0.5$ ps. This yields a theoretical speedup factor of 20 and 50 over MD, respectively.

A plot of the coherent term $(a_1 - a_0 b_1)$ in Eq. (5) as a function of time is shown in Fig. (5). This term evolves with relatively small fluctuations in time, as expected, and it provides an accurate approximation of the velocity of the CG field variable, $\Pi$. This is important particularly for problems in which the inertial regime dominates; in this case, $\Pi$ must be incorporated as part of the CG description. In contrast, the $a_{1/2}$ term in Eq. (4) fluctuates dramatically from one CG timestep to another as shown in Fig. (6). This confirms the motivation for guiding the construction of the TPA via the Ito formula. It also illustrates the notion that analysis of the TPA naturally allows for the separation of coherent versus stochastic aspects of macromolecular systems. Furthermore, the ratio of noise to coherence factors in Eq. (4) is shown in Fig. (7), which shows this ratio can be used as a metric for deciding if a particular coarse-grained variable is slowly varying in time. In this case, this ratio is small, which implies $\phi_m$ is suitable for multiscale simulation. Finally, the CG timestep $\Delta$ can be varied in time using the TPA formula as covered in the appendix. Fig. (8) shows how the parameter $\alpha$, which scales $\Delta$ for ever CG timestep, can be controlled based on estimating the local error that results from the TPA scheme.

### Protein migration

Long time stability of the TPA scheme was assessed by simulating an HPV L1 protein (PDB ID: 1DZL) in aqueous solution at 310 K and subjected to an externally forced flow. The force used to induce the flow was applied on all water molecules in the $z-$direction, as detailed in. The external force had a magnitude of 0.1 kJ/mol nm, which is approximately equivalent to a pressure gradient of 54.6 bar/nm. The momentum field variable chosen for
this problem is
\[ \phi_{p,z}(r, s_l) = \sum_{i=1}^{N} p_{i,z} K(r_i - s_l, w), \] (18)
with the width \( w \) set to 30 Å. This CG variable was used to capture the migration of the protein through a graphene nanopore, which is modeled as a set of static alpha carbon atoms interacting via VdW forces (Fig. (9)). NaCl counter-ions of concentration 0.15 M were added for charge neutrality. The system consisted of 3,721,309 atoms in a box of dimensions [0,30.7] nm, [0,30.27] nm, and [0,40] nm along the \( x, y, \) and \( z \) axes, respectively. The center of mass of the protein was located at [14.06, 15.51, 5.07] nm after energy minimization was performed on the system (Fig. (10)). An equilibration run with position restraints imposed on the protein was performed for 100 ps; after thermal equilibrium was established, the system was simulated under NVT conditions for 1 ns during which the protein moved through the nanopore.

A grid comprising \( 4 \times 4 \times 4 \) nodal points was used to capture the CG features of the HPV L1 protein migration driven by an externally forced flow. This yielded a total of 64 field variables. To illustrate the dynamics, the flow (characterized by the momentum density \( \phi_{p,z} \) at \( s_l = [13.46, 13.56, 10.05] \) nm) was found to be smoothly varying as shown in Fig. (11). The \( z \) component of the momentum field variable increases almost monotonically in time due to the constant force applied along the \( z \)-axis. The TPA scheme accurately reproduces the temporal profile of \( \phi_{p,z} \) for \( \Delta = 5 \) ps yielding a theoretical speedup of 5 over MD. As \( \Delta \) is increased to 15 ps (yielding a theoretical speedup of 15 over MD), \( \phi_{p,z} \) computed by TPA initially experiences significant fluctuations that can be reduced by using adaptive time stepping. This shows the TPA scheme has the potential of reproducing the temporal profile of stochastic variables over long periods of time and with minimal error.
Conclusion

An implicit time integration scheme based on Padé approximants was introduced for solving differential equations that arise in multiscale molecular dynamics. The scheme provided improved accuracy and stability over the projective forward Euler method used earlier,\textsuperscript{2,3} and a natural way to separate the stochastic from the coherent components of the evolution; this yields physical insights into problems such as structural transitions in supramolecular assemblies, and possible critical fluctuations associated with them.

The time integration scheme introduced here can be extended to other problems such as porous media flow, wherein the microphase is the grain scale and the CG description is for much larger scales, and simulation techniques such as dissipative particle dynamics\textsuperscript{32} and multiscale Langevin methods.\textsuperscript{33}

Acknowledgements

This work was supported by in part by the NSF division of material science research (grant 1533988, 80\%) and the Indiana Clinical and Translational Sciences Institute, funded in part by grant UL1 TR001108 from the National Institutes of Health, National Center for Advancing Translational Sciences, Clinical and Translational Sciences Award (20\%). The authors also acknowledge support by the IU College of Arts and Sciences via the Center for Theoretical and Computational Nanoscience, and Lilly Endowment, Inc., through its support for the Indiana University Pervasive Technology Institute, and in part by the Indiana METACyt Initiative. The Indiana METACyt Initiative at IU is also supported in part by Lilly Endowment, Inc. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.
Appendix

Using a variable order method, the CG timestep $\Delta$ in Eq. (4) can be varied by estimating and controlling some measure of the global error. For example, performing second and third order Maclaurin expansions in terms of $\tau$ for the exponential term in Eq. (4) yields the following truncated TPAs:

$$\phi^{(2)}(\tau) = -\frac{1}{2}(a_0 + a_{1/2}^{0.5} + a_1 \tau)(2 - 2b_1 \tau + b_1^2 \tau^2) + O(\tau^3), \quad (19)$$

$$\phi^{(3)}(\tau) = -\frac{1}{6}(a_0 + a_{1/2}^{0.5} + a_1 \tau)(-6 + 6b_1 \tau) - 3b_1^2 \tau^2 + b_1^3 \tau^3 + O(\tau^4). \quad (20)$$

The local error that results from the TPA scheme is determined by taking the difference between Eq. (20) and (19); the new timestep $\Delta$ that keeps the local error less than or equal to $tol$ is then estimated to be

$$\Delta \leq \frac{1}{b_1} \left( \frac{6 \times tol}{a_0} \right)^{\frac{1}{3}} \times \Delta_p, \quad (21)$$

where $\Delta_p$ is the previous timestep. The fraction $\alpha$ that controls the adaptive time stepping (Fig. (8)) is therefore $\frac{1}{b_1} \left( \frac{6 \times tol}{a_0} \right)^{\frac{1}{3}}$.

References


(4) Mansour, A. A.; Sereda, Y. V.; Yang, J.; Ortoleva, P. J. Vaccine 2015, 33, 5890–5896.


\[
\begin{align*}
\phi(t - \Delta) & \quad \phi(t - \Delta + \delta) & \quad \phi(t) & \quad \phi(t + \delta) & \quad \phi(t + \Delta) \\
\end{align*}
\]

Figure 1: The TPA scheme predicts the future state of \( \phi \) at \( t + \Delta \) using historical information at \( t - \Delta, t - \Delta + \delta, t, \) and \( t + \delta \).

Figure 2: Pertussis toxin protein (PDB ID: 1PRT) undergoes a conformational change (A) under NVT conditions at \( T = 300 \) K. The RMSD of the protein (B) increases almost monotonically over a period of 2 ns.
Figure 3: A plot of the average % relative error \( \langle |\phi_m - \phi_m^{\text{est}}|/\phi \times 100 \rangle \) where \( \phi_m^{\text{est}} \) is estimated using MF or MFI, and \( \phi_m \) extracted from MD, shows that for \( \Delta = 0.2 \) ps, this error increases in time for MF whereas it remains bounded for MFI.

Figure 4: For \( \Delta = 0.2 \) ps, MFI accurately reproduces the trajectory of the local mass density, \( \phi_m \), of pertussis toxin as the protein undergoes compression-extension. As \( \Delta \) is increased to 0.5 ps, accuracy decreases as expected, but stability is maintained.
Figure 5: The CG velocity $\Pi$ obtained from MD (black) is compared to that computed using (a) forward finite difference (teal), which amplifies the noise in $\phi$ thus leading to inaccuracies in $\Pi$ and (b) the truncated TPA in Eq. (5) (blue), which shows the TPA scheme is accurate in unraveling the coherent from the stochastic effects.

Figure 6: The stochastic term $a_{1/2}$ fluctuates rapidly in time, with its time average close to zero, as suggested by the Ito formula used to construct the TPA in Eq. (4).
Figure 7: The dimensionless ratio of noise ($a_{1/2}$) to coherence ($\left((a_1 - a_0 b_1)\Delta^{1/2}\right)$ factors in Eq. (5) suggests the mass density $\phi_m$ used here is slowly varying in time and therefore serves as a basis for multiscale factorization.

Figure 8: The CG timestep $\Delta$ can be varied in time to control accuracy and reduce truncation errors that result from TPA. The $\alpha$ factor is used to scale the previous CG timestep based on Eq. (21) shown in the appendix.
Figure 9: An HPV L1 protein migrates through a static graphene nanopore (of radius 7.5 nm) due to an externally forced flow along the z-axis.
Figure 10: A snapshot of the HPV L1 protein in aqueous solution at $t = 0$ ns. The protein moves through the graphene nanopore as directed by an external force applied in the z-direction.
Figure 11: The momentum density along the z direction is generally coherent and decreases almost monotonically in time as the protein migrates towards the graphene nanopore. MFI captures this migration by reproducing the trajectory of $\phi_{p,z}$ for $\Delta = 5$ ps and 15 ps.
Graphical TOC Entry

\[ \phi^{CA}(t - \Delta t) = a_0 + a_1 \tau^{1/2} + a_2 \tau \]