International Journal of Non-Linear Mechanics 43 (2008) 1040 – 1055

Contents lists available at ScienceDirect

International Journal of Non-Linear Mechanics

journal homepage: www.elsevier.com/locate/nlm

Substructured molecular dynamics using multibody dynamics algorithms

Rudranarayan M. Mukherjee[a,*,1], Paul S. Crozier[b], Steven J. Plimpton[b], Kurt S. Anderson[a]

[a]Rensselaer Polytechnic Institute, 110 8th Street, Troy NY 12180, USA
[b]Sandia National Laboratories, Albuquerque, New Mexico, USA

ARTICLE INFO

Article history:
Received 31 October 2007
Received in revised form 1 April 2008
Accepted 2 April 2008

Keywords:
Coarse-grained molecular dynamics
Multibody dynamics algorithms
Open-source software

ABSTRACT

This paper reports a new research effort aimed at using efficient multibody dynamics methods to simulate coarse-grained molecular systems. Various molecular systems are studied and the results of nanosecond-long simulations are analyzed to validate the method. The systems studied include bulk water, alkane chains, alanine dipeptide and carboxyl terminal fragments of calmodulin, ribosomal L7/L12 and rhodopsin proteins. The stability and validity of the simulations are studied through conservation of energy, thermodynamics properties and conformational analysis. In these simulations, a speed up of an order of magnitude is realized for conservative error bounds with a fixed timestep integration scheme. A discussion is presented on the open-source software developed to facilitate future research using multibody dynamics with molecular dynamics.

Published by Elsevier Ltd.

1. Introduction

Simulations of the dynamics of biomolecular and polymeric systems are usually characterized by very complex energy surfaces and large conformational spaces. Consequently, there are two fundamental computational challenges which have limited the scope of molecular dynamics (MD) simulations. The first challenge is associated with the spatial size of these systems which makes the calculation of interactions a computationally expensive step in the simulations. The biomolecular or polymeric systems of interest typically contain large number of interaction sites (or atoms) and these systems can span hundreds to many thousand degrees of freedom (n).

The calculation of the interaction forces between these sites is a many body problem which traditionally scales as $O(n^2)$. Methods such as [1–3] have been proposed that reduce the cost of this calculation. Although these methods reduce the cost of the force calculations, considering the large value of n, this calculation is still computationally expensive and it accounts for most of the computational cost of a simulation.

The second challenge is associated with the diverse frequency content of the system and the large temporal domains that contain the phenomena of interest. As the integration step size is inversely proportional to the highest frequency content of the system, a MD simulation of a fine scale system typically propagates with an integration timestep of only subfemtoseconds. Consequently, the state-of-the-art *long* simulation ranges from nanoseconds to micro-seconds of real time. Many phenomena of interest occur in longer temporal scales ranging from milliseconds to several seconds.

The methods aimed at reducing the computational complexity of MD can be broadly classified into two classes. The first class of methods are those that aim at reducing the complexity of calculating the interaction forces between different elements in a system by re-parameterizing the nature of the force fields [3–6]. These methods have typically focused on producing estimates of the forcing terms, with acceptable accuracy, at greatly reduced cost. The second class of methods is aimed at constraint imposition to truncate the frequency spectrum of the system to allow larger temporal integration step sizes. The different methods that fall under this class include the following.

Constrained dynamics using explicit constraints [7–10]: The explicit constraints are coupled with the equations of the unconstrained system resulting in a set of *d*ifferential-*a*lgebraic *e*quations (DAEs) (differential equations of motion and algebraic constraint equations). While representative methods such as SHAKE [7], and others [8–10] have been successfully used, the cost and numerical issues involved in iteratively solving the DAEs can be challenging for larger systems and different types of constraints.

Constrained dynamics using implicit constraints [11–13]: This family of methods is commonly called internal coordinate MD. They satisfy the constraints exactly and implicitly through the use of generalized coordinates. The solution of the heavily coupled equations

* Corresponding author. Tel.: +1 818 354 2677.
E-mail addresses: Rudranarayan.M.Mukherjee@jpl.nasa.gov (R.M. Mukherjee), pscrozi@sandia.gov (P.S. Crozier), sjplimp@sandia.gov (S.J. Plimpton), anderk5@rpi.edu (K.S. Anderson).
[1] Now at NASA Jet Propulsion Laboratory, Pasadena, CA.

0020-7462/$ - see front matter Published by Elsevier Ltd.
doi:10.1016/j.ijnonlinmec.2008.04.003

of motion can be computationally expensive unless intelligent, efficient schemes are employed. Reduced computational order algorithms [15–18] have been proposed to reduce the cost of these methods. They use special manipulations to decompose and solve the equations of motion in particular efficient ways (at least for systems involving many generalized coordinates), thus avoiding much of the very costly linear algebra.

Multirate temporal integration [18–24]: These methods exploit the multiscale character of biopolymer motions. In what is known as multiple time-scale (MTS) schemes [25] a hierarchy of timesteps is introduced instead of the traditional single timestep. The idea is to take advantage of the terms in the biomolecular Hamiltonian that vary more slowly than other terms. Appropriate timesteps are assigned for the different terms.

Eigenvector (*modal*)*-schemes* [26–31]: These schemes use normal modes (free vibration modes), which are obtained from the diagonalization of the Hessian matrix (second derivative of the potential). These modes are used to describe global deformations around a local minimum on the molecular potential energy surface. While these methods allow for a potentially large integration step size, the associated linearization and assumption of modal superposition can restrict the efficacy of these methods.

Implicit integration schemes [32–34]: Implicit integration schemes are interpolative in nature, and thus may take much larger stable timesteps than explicit integration schemes which are extrapolative. Compared to explicit integration methods, implicit integration schemes are more expensive as they usually require the solution of a set of coupled algebraic equations. These non-linear systems of equations may be expensive to solve, thereby offsetting any gains made by larger timesteps. Further, the damping introduced by the implicit schemes may pose a problem in maintaining the micro-canonical conditions of the ensembles.

In this paper, a methodology is presented for coarse-grained molecular dynamics (CGMD) simulations by selecting and combining several existing algorithms from different fields of application. Coarse-grained models of molecular systems are developed by imposing constraints that curtail the frequency spectrum, thereby allowing for a larger integration step size. The coarse-grained models of the systems are in the form of articulated multirigid body systems. This method uses efficient multibody dynamics algorithms for solving the equations of motion of constrained MD systems in linear complexity. It uses generalized coordinates and imposes the holonomic constraint implicitly and exactly, not just to integrator tolerance. A two-step iterative integration scheme is used for temporally advancing the states. This scheme requires only one system wide force calculation and takes into account the velocity-dependent gyroscopic and Coriollis forces. The method uses Cartesian force fields and gains from the computational savings obtained from using efficient force field calculation schemes such as Ewald sum and the P^3M methods [1,2]. The use of efficient multibody dynamic algorithms coupled with efficient and highly parallelizable force field calculations is a novel research effort and results in a highly efficient computation method for constrained MD. Two different schemes for constraint imposition are used, viz holonomic distance (and/or) angle constraints as well as aggregation. In the case of holonomic distance constraints, the atomistic representation of the system is retained. However, in the aggregation model, the system is coarse grained into clusters, where each cluster may contain more than one atom. This method is implemented in a C++ object oriented open-source software developed by coupling together the highly parallel MD software LAMMPS (Large-Scale Atomic/Molecular Massively Parallel Simulator) [35] and the efficient multibody software POEMS (Parallelizable open-source Efficient Multibody Software) [36]. The original contributions of this manuscript are (i) the development and validation of the multibody dynamics based method for CGMD simulations and (ii) the development of a research code that embodies this method.

2. Method overview

A number of efficient algorithms and models have been synergistically combined together to form a method to facilitate substructured MD simulations. In this method two key aspects, efficient calculation of molecular interaction and efficient solution of multibody equations of motion, have been seamlessly integrated with various aspects of MD simulation protocols. The method supports a whole host of interaction potential fields used for modelling molecular systems. Similarly, the method builds on efficient multibody algorithms such as the $O(n)$ method [43] and the generalized impulse momentum algorithm [45]. The systems studied are coarse grained into articulated system of rigid bodies. As evidenced from the various systems studied, this coarse graining eliminates the high frequency dynamics of the system thereby allowing for a larger integration timestep. This results in significant reduction in the overall computational cost of the simulations.

In this method, the systems of interest are coarse grained into different hierarchies of reduced order models that are subject to implicit constraints. The reduced order models and the constraints eliminate the high frequency components of the system dynamics and allow for a stable, large integration timestep. There are two different representations of the reduced order models, both of which are in the form of an articulated multibody system. The kinematic model used here enforces the constraints implicitly, exactly and non-iteratively using the topological information and generalized (relative) coordinates, thereby reducing the computational complexity.

The first category of models is obtained by imposing holonomic distance-dependent constraints between the interaction sites. The fully atomistic representation of these systems consists of interaction sites connected to each other by stiff springs. In the reduced order model, the axial stiff spring is replaced by a constant length massless rigid link. Each link and the next interaction site it is attached to is treated as a single body with a point mass and negligible moments of inertia. Successive bodies are connected to their inward bodies by kinematic joints allowing only rotational degrees of freedom.

The second category of models is obtained by *aggregation* where a large number of discrete interaction sites are constrained to move as an articulated system of coupled rigid bodies. Each rigid body consists of a number of interaction sites that are coupled together to move as a rigid body. Two rigid bodies are coupled when they share a common interaction site. This common interaction site is treated as a kinematic joint location. The relative motion between two rigid bodies is modelled using joints that allow relative rotational motion and constrain out the translational motion. The relative rotation between bodies in an articulated model is modelled using Euler parameters to avoid any singular configurations. The dynamics of these reduced order models is simulated using the efficient multibody dynamics algorithm described in Section 3.

Among the systems studied using this method, the periodic boundary condition has been used extensively in the simulation of explicit water, alkanes and other polymer simulations. The protein simulations are carried out using the *shrink-wrap* boundary conditions where the simulation box is non-periodic and the position of the face is set so as to encompass the atoms in that dimension (shrink-wrapping), no matter how far they move.

In this method, the system description is concurrently maintained using two sets of coordinates viz generalized coordinates and Cartesian coordinates. The system is represented in generalized coordinates to describe the kinematics and dynamics of the reduced order models. The generalized coordinates describe the relative orientation and displacement of the reduced order models. The underlying multibody dynamics algorithm also generates the fully atomistic representation of the system in Cartesian coordinates at