Structure, Vol. 9, R45–R53, April, 2001, ©2001 Elsevier Science Ltd. All rights reserved. PII S0969-2126(01)00593-7

Time-Trimming Tricks for Dynamic Simulations: Splitting Force Updates to Reduce Computational Work

Ways & Means

Tamar Schlick*
Department of Chemistry and Courant Institute
of Mathematical Sciences
New York University and the Howard Hughes
Medical Institute
251 Mercer Street
New York, New York 10012

Introduction

As one of the largest group of supercomputer consumers, macromolecular modelers are always scrambling for more computer time. With the goal of gaining insights into a variety of biological processes, scientists are also continuously trimming down algorithmic cost. This economizing of code performance time while maximizing the resulting biological information can be pursued by using sophisticated mathematical and computer science machinery where available (e.g., for fast electrostatic summations [1] or nonbonded-list manipulations), by approximating where possible (e.g., continuum solvation [2], conformational path estimates [3]), or any combination of the above.

Though a variety of simulation techniques are available (see Figure 1), the molecular dynamics (MD) approach tops the list in popularity because of its physical appeal and biological connection. Namely, the trajectories showing the evolution in time and space of molecular conformations follow classical physics; this allows kinetic processes to be followed in detail, linking and expanding upon experimental observations. MD's popularity would be overwhelming if the technique's computational demands, and hence biological scope, were not so limited for large systems. This limitation stems from the numerical stability requirement, which restricts the timestep size used for integrating the equations of motion to a relatively small value (e.g., $\sim$1 fs). This timestep size implies *one million* or more steps for simulating a mere nanosecond in the life of a biomolecule; this number already translates to days or months of computing time, even on state-of-the-art platforms [4], since each step typically requires at least one expensive force evaluation. (In empirical molecular force fields, the force is defined as the negative gradient of the total energy [enthalpy], which is expressed as a sum of harmonic bond length and bond angle expressions, trigonometric torsion terms, nonbonded Coulomb and van der Waals components, and other terms.)

Certainly, code parallelization on multiple-processor machines shaves off computing clock time, as demonstrated by the longest simulation to date, 1 μs, for a villin headpiece, achieved in 4 months of dedicated CPU time on 256 processors of a Cray T3D/E [5], or by IBM's ambitious announcement to fold proteins by 2005 on a unique petaflop computer called Blue Gene with a massively parallel architecture (one million or more processors). Any algorithmic speedup will only help to further reduce MD computational time.

One of the golden opportunities for work reduction is in the time integration protocol. Since the overall cost of MD simulations is dominated by the number of force evaluations, reducing this number—typically by increasing the effective timestep size—can lengthen the physical time span that can be simulated. Yet, this "timestep problem" [6, 7] has remained a tough nut to crack. Much of this difficulty can be simply attributed to the dilemma of balancing a larger step size with lower resolution accuracy. That is, with decreased intervals of molecular observations (force update frequency), the resolution of some fast processes might have to be sacrificed. Another obstacle is practical: the more sophisticated algorithms, such as those described below, are not trivial to incorporate in the context of large programs, since evaluations of the total force depend on many program units. This makes the application of new integrators for MD more complex than fast schemes for electrostatics summations.

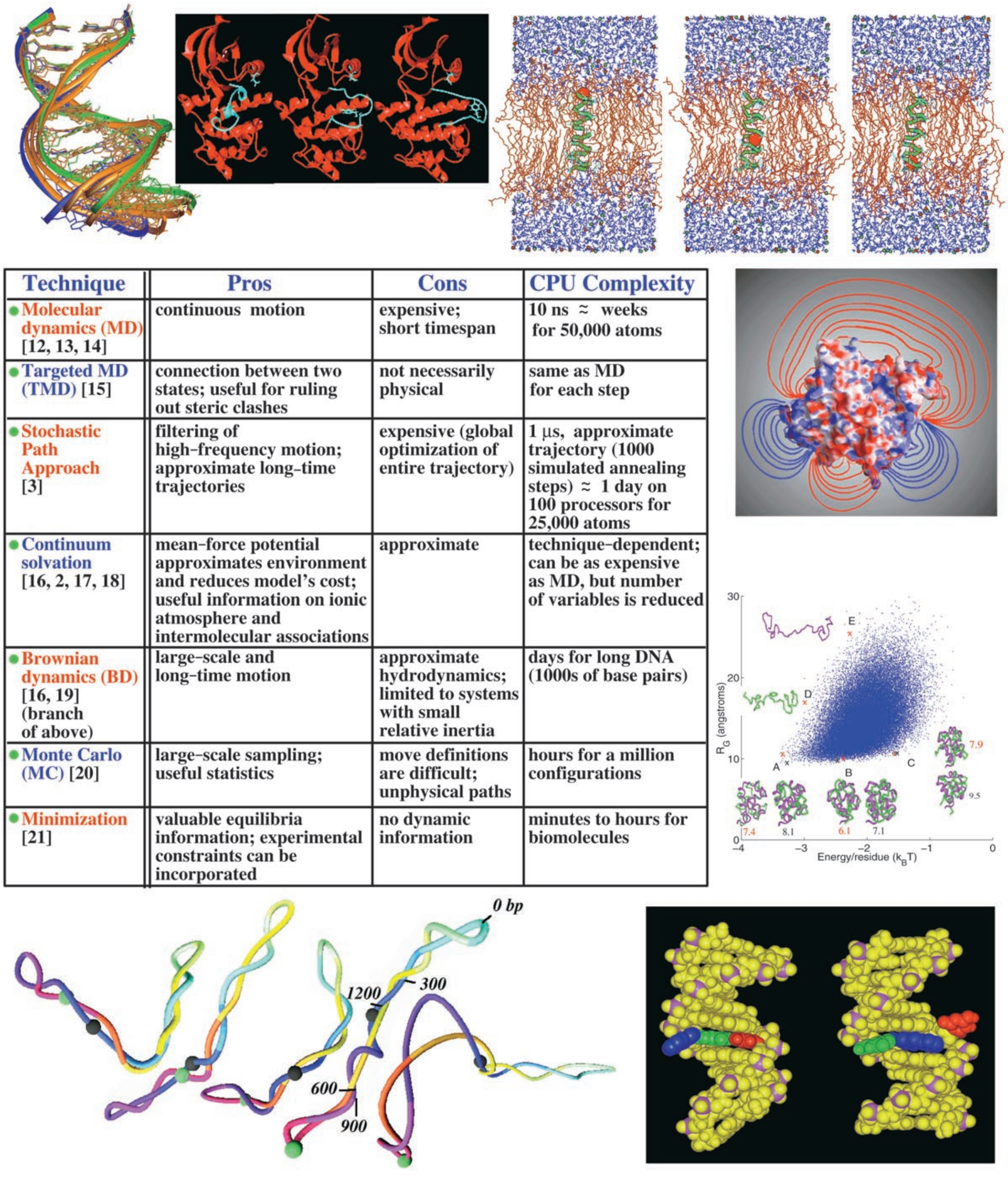
In the evolution of certain physical systems—motion of planets or flow of compound concentrations in chemical reactions—the effect of the fast processes on the global motion is negligible. In biomolecular systems, unfortunately for algorithm developers, vibrational modes are intricately coupled: thus, fast small-amplitude motions can trigger a cascade of events that culminate in large-scale global rearrangements. This intricate coupling limits the traditional mathematical machinery available for MD integration and compels algorithm developers to seek inventive, tailored approaches (see Figure 1). Still, against these challenges, mathematicians and other computational scientists have labored in the past decade to understand the numerical limitations of MD integration, devise clever schemes, and design approaches departing from accurate motion-following that instead yield greater overall information on the large thermally accessible configuration space of macromolecules [8, 7, 9, 10].

Following a historical perspective of method development, we sketch the mathematical machinery for the promising integration approach termed "force splitting" or "multiple timesteps" (MTS). We describe the associated difficulties in standard MTS implementations (resonance artifacts), outline how they can be overcome at the cost of augmenting the underlying Hamiltonian (as in the LN approach [6, 11]), and illustrate how these methods can be evaluated and used to study large biological systems. We conclude by outlining some outstanding problems that remain, such as the efficient implementation of MTS schemes in combination with Ewald summations and in applications to various thermodynamic ensembles.

Method Development: Historical Perspective

Since the 1960s, the Verlet/Störmer method [32, 28] has been the gold standard for MD (see Figure 2 for a discret-

*To whom correspondence should be addressed (e-mail: schlick@nyu.edu).



Technique	Pros	Cons	CPU Complexity
● Molecular dynamics (MD) [12, 13, 14]	continuous motion	expensive; short timespan	10 ns ≈ weeks for 50,000 atoms
● Targeted MD (TMD) [15]	connection between two states; useful for ruling out steric clashes	not necessarily physical	same as MD for each step
● Stochastic Path Approach [3]	filtering of high-frequency motion; approximate long-time trajectories	expensive (global optimization of entire trajectory)	1 μs, approximate trajectory (1000 simulated annealing steps) ≈ 1 day on 100 processors for 25,000 atoms
● Continuum solvation [16, 2, 17, 18]	mean-force potential approximates environment and reduces model's cost; useful information on ionic atmosphere and intermolecular associations	approximate	technique-dependent; can be as expensive as MD, but number of variables is reduced
● Brownian dynamics (BD) [16, 19] (branch of above)	large-scale and long-time motion	approximate hydrodynamics; limited to systems with small relative inertia	days for long DNA (1000s of base pairs)
● Monte Carlo (MC) [20]	large-scale sampling; useful statistics	move definitions are difficult; unphysical paths	hours for a million configurations
● Minimization [21]	valuable equilibria information; experimental constraints can be incorporated	no dynamic information	minutes to hours for biomolecules

Figure 1. Various Simulation Approaches

Continuum solvation includes empirical constructs, generalized Born models, stochastic dynamics, or Poisson Boltzmann solutions. CPU times are rough estimates for single processors (e.g., SGI Origin 2000 300 MHz R12000) unless stated otherwise. Molecular images surrounding the table illustrate the various techniques. From top left, clockwise: MD snapshots of DNA TATA element showing the captured bending toward the major groove [22]; TMD snapshots of the Hck protein, where the restrained segment (blue) moves towards the conformation found in the active form of the enzyme (M. Young and J. Kuriyan [23]); stochastic-path approach snapshots describing the permeation of a sodium ion (huge red sphere) through the gramicidin channel embedded in a DMPC membrane (K. Siva and R. Elber, personal communication); electrostatic potential contours of mouse acetylcholinesterase (D. Sept, K. Tai, and J.A. McCammon, personal communication) [24]; MC configurations and ensemble radius of gyration/energy plots in a folding simulation for 434 repressor protein based on statistical potentials (RMS values given for superimposed predictions with native structure) [25]; NMR solution structures, delineated using minimization subject to NMR constraints, of a DNA duplex with a carcinogen 2-aminofluorene DNA adduct that adopts two conformations in equilibrium [26]; and BD snapshots of 1500 bp DNA showing site-juxtaposition kinetics at high salt (0.02 M monovalent ions) of two segments located 500 bp along the DNA contour [27].

Verlet: The Gold Integration Standard

We write the Newtonian equations of motion for a system of N atoms as as the following pair of first-order differential equations:

$$\mathbf{M}\dot{V}(t) = F(X) = -\nabla E(X(t)) + \dots , \dot{X}(t) = V(t)\,, \quad (1)$$

where X and V are the collective vectors of position and velocity ($3N$ components), $\mathbf{M}$ is the diagonal mass matrix, and the dot superscripts denote differentiation with respect to time, t. The total force F is composed of the systematic force, which is the negative gradient of the potential energy E, and possibly additional terms. For example, in the simple Langevin model, these additional terms represent a frictional force proportional to the velocity, and random Gaussian force with given statistical properties:

$$\mathbf{M}\dot{V}(t) = -\nabla E(X(t)) - \gamma \mathbf{M}\dot{X}(t) + R(t), \quad (2)$$

$$\langle R(t) \rangle = 0, \quad \langle R(t)R(t')^T \rangle \; = \; 2\gamma k_B \mathrm{T}\mathbf{M}\, \delta(t-t'). \quad (3)$$

Here γ is the damping constant (in reciprocal units of time), k_B is the Boltzmann constant, T is the target temperature, and δ is the Dirac delta function. Together, these additional terms mimic a thermal reservoir at temperature T.

The original Verlet algorithm [29] describes trajectory positions for $n = 1, 2, \cdots$, at time intervals Δt from a given X^0 by

$$X^{n+1} = 2X^n - X^{n-1} + \Delta t^2 F^n\,, \quad (4)$$

where the superscripts n denote the finite-difference approximations to quantities at time $n\Delta t$. The flexibility in defining compatible velocity formulas from this Verlet position propagation formula has led to Verlet variants. Though sharing theoretical properties, numerical properties of these variants may be different due to computer roundoff error.

Three popular variants of Verlet are the *leapfrog* [31], *velocity Verlet* [32], and *position Verlet* schemes. The force is evaluated at the endpoints in leapfrog and velocity Verlet but at the midpoint of position Verlet.

The leapfrog and velocity Verlet schemes are defined, respectively, as:

$$\begin{aligned} V^{n+\frac{1}{2}} &= V^{n-\frac{1}{2}} + \Delta t\, F^n \\ X^{n+1} &= X^n + \Delta t\, V^{n+\frac{1}{2}}\,, \end{aligned} \quad (5)$$

and

$$\begin{aligned} X^{n+1} &= X^n + \Delta t\, V^n + \frac{\Delta t^2}{2} F^n \\ V^{n+1} &= V^n + \frac{\Delta t}{2}\,(F^n + F^{n+1})\,. \end{aligned} \quad (6)$$

Both variants can be written in a symmetric fashion using the following *propagation triplet*:

$$\begin{aligned} V^{n+\frac{1}{2}} &= V^n + \frac{\Delta t}{2} F^n \\ X^{n+1} &= X^n + \Delta t\; V^{n+\frac{1}{2}} \\ V^{n+1} &= V^{n+\frac{1}{2}} + \frac{\Delta t}{2} F^{n+1}\,. \end{aligned} \quad (7)$$

In the popular extension of Verlet to Langevin dynamics [33], each force expression F^n above is simply replaced by $F^n - \gamma V^n + \mathbf{M}^{-1} R^n$.

Another variant of Verlet is known as *position Verlet*, written as the triplet

$$\begin{aligned} X^{n+\frac{1}{2}} &= X^n + \frac{\Delta t}{2} V^n \\ V^{n+1} &= V^n + \Delta t\, F^{n+\frac{1}{2}} \\ X^{n+1} &= X^{n+\frac{1}{2}} + \frac{\Delta t}{2} V^{n+1}\,. \end{aligned} \quad (8)$$

Figure 2. Integration Framework for the Verlet Method

ization outline). The Verlet method forms the basis of virtually all extensions used today, such as for constrained dynamics, Langevin dynamics, MTS methods, and various thermodynamic ensembles. Verlet's good behavior can be attributed to its *symplecticness*, a favorable numerical property for conservative Hamiltonian systems that roughly implies preservation of volumes in phase space [33]. Essentially, this also means that the simulated trajectory lies sufficiently close to that corresponding to the exact, governing Hamiltonian.

In an attempt to reduce MD computational times, Verlet-based approaches for constraining the fastest degrees of freedom (e.g., bond lengths) soon became popular (e.g., "SHAKE" and its variants) [34, 35]. Such approaches increase the timestep slightly (e.g., to around 2 fs), with modest added cost [34, 35]. Extensions for freezing the next fastest vibrational mode (heavy atom angle bending) fail due to strong vibrational coupling [36, 37]. In a similar spirit, internal-coordinate MD approaches were attempted (e.g., with torsion angles, rather than Cartesian coordinates, as the independent variables) [38–40]. While certainly reducing the number of variables, these schemes complicate the propagation expressions and, more seriously, can slow down barrier-crossing events that are facilitated by local motions. Still, internal coordinate MD approaches have found good usage in structure refinement [41].

In the 1970s, *multiple-timestep* (MTS) variants were introduced to reduce the computational cost of MD simulations [42]. MTS methods represent a special case of multilevel techniques in applied mathematics [43]. They are based on the rather simple idea of *force splitting*: update the *slowly varying* forces *less often* than the rapidly varying terms. Savings can be realized if the slowly varying forces due to distant interactions (e.g., electrostatics) are held constant over longer intervals than the more rapidly varying short-range forces (e.g., bonded terms). Standard Verlet procedures can then be modified by evaluating the long-range forces less often than the short-range terms (see Figure 3). Between updates, the slow forces can be incorporated into the discretization either as constants (via *extrapolation*) or as zero forces (i.e., via *impulses*, that is, considered only at the beginning and end of every macrostep). As illus-

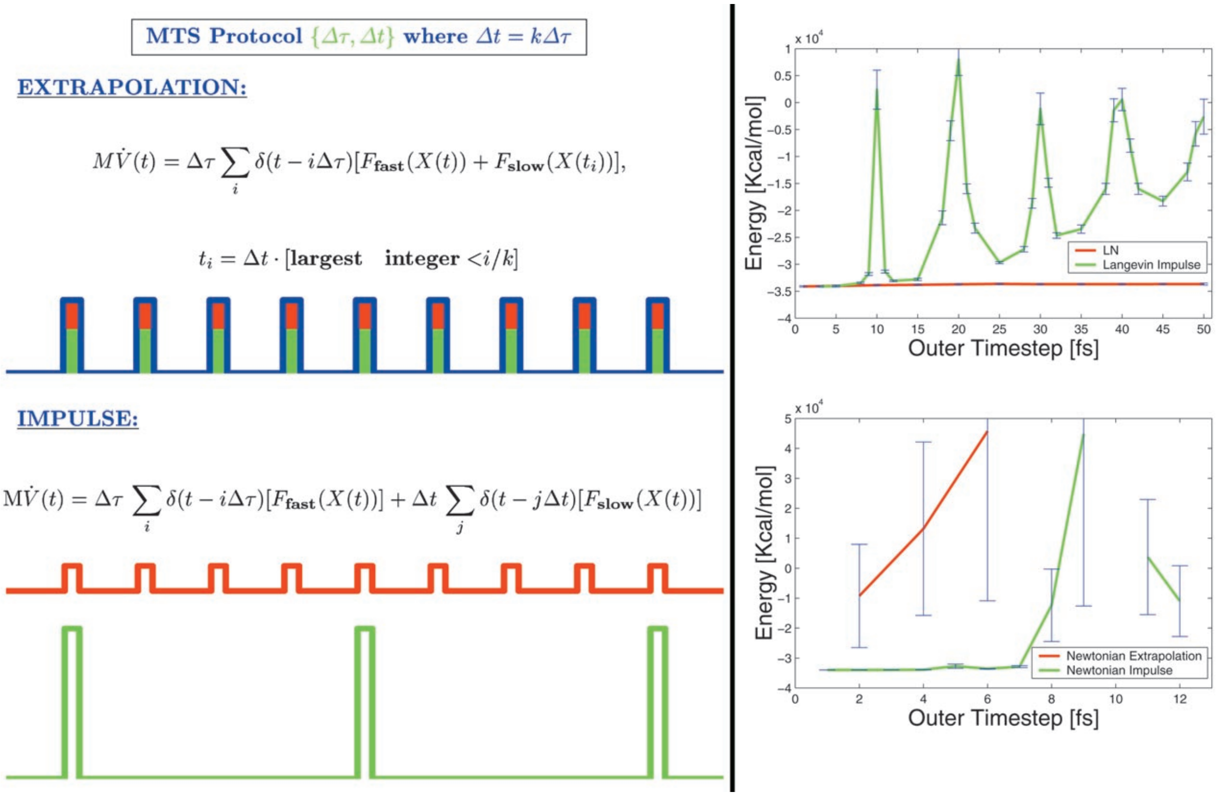


Figure 3. Extrapolative versus Impulse Force Splitting

(Left) Schematic illustration for a dual-timestep protocol with inner timestep $\Delta\tau$ and outer timestep $\Delta t = k\ \Delta\tau$ ($k = 4$ used). In extrapolative splitting, a slow-force contribution (green) is made each time the fast force (red) is evaluated (i.e., every $\Delta\tau$ interval). In impulse splitting, in contrast, contributions of the slow forces (tall spikes) are considered only at the time of their evaluation (e.g., every four fast-force evaluations). (Right) Examples of extrapolation and impulse applications to Newtonian and Langevin dynamics for a solvated BPTI (bovine pancreatic typsin inhibitor) system. Resonance artifacts for impulse splitting are apparent for both Newtonian and Langevin dynamics, as are the energy drift for Newtonian/extrapolation and the stability of Langevin/extrapolatition.

trated in Figure 3 for a two-timestep scheme, the slow forces contribute to each small timestep in extrapolation, even though they are only computed every long timestep, rather than being added only at their time of evaluation, as in impulse splitting.

For example, consider a three-timestep protocol ($\{\Delta\tau$, Δt_m, $\Delta t\}$), where the timesteps are related by integral multiples: the small timestep $\Delta\tau$ defines the update frequency of the local, bonded forces; the medium timestep Δt_m, an integer multiple of $\Delta\tau$ ($\Delta t_m = k_1\ \Delta\tau$), specifies a longer update frequency for nonbonded interactions within a given radial distance (e.g., 6 Å region); and an outermost (large) timestep, $\Delta t = k_2\ \Delta t_m$, defines the frequency of calculating the remaining, long-range forces. By splitting the forces appropriately and maximizing the outer timestep (or the ratio $r = \Delta t/\Delta\tau = k_1\ k_2$ between the largest to smallest timestep), we can optimize the overall *speedup* in relation to a single-timestep (STS) simulation at $\Delta\tau$. This partitioning reduces computational time, since the majority of the work in the total force evaluation stems from the slow components [11]. These long-range terms are order $\mathbb{O}(N^2)$ in direct complexity (or $\mathbb{O}(N \log N)$ by accelerated methods [1]), in contrast to linear complexity for the local terms (N is the number of independent variables, typically atoms). As discussed below, when particle mesh Ewald (PME) methods are used for the electrostatic calculations, splitting the PME direct and reciprocal terms [1] into force classes remains a challenge, since the reciprocal component contains rapidly varying terms.

While intuitive and reasonable, such MTS approaches were not immediately as successful and simple as had been hoped. The details of merging the updates of the different force components required great care to avoid instabilities. More significantly, attempts to increase the outer timesteps encountered unexpected, and theoretically peculiar, obstacles; namely, even though the fast forces were updated frequently, a much larger timestep was not possible as might appear appropriate for the slow components.

Specifically, the early extrapolative MTS variants exhibited systematic energy drifts [44] and were abandoned. Symplectic variants followed [45, 46], with important mathematical machinery presented for their analysis and development (Trotter factorization of the Liouville operator) [46]. However, early observations [45, 47] forecasted hidden dangers in these symplectic impulse-MTS treatments. Indeed, in 1995, when these methods were applied to biomolecular dynamics [48, 49], a limitation on the outermost timestep to less than 5 fs was observed. This value not only limited practical speedup, it posed a puzzle for its origin, given that 5 fs

```
Extrapolative MTS
based on Position Verlet

$X_r^0 \equiv X + \frac{\Delta t_m}{2} V$                                  (1a)
$F_{\mathrm{slow}} \equiv -\mathbf{M}^{-1} \nabla E_{\mathrm{slow}}(X_r)$  (1b)

For $j = 1$ to $k_2$
    $X_r \equiv X_r^j \leftarrow X + \frac{\Delta t_m}{2} V$               (2a)
    $F_{\mathrm{med}} \equiv -\mathbf{M}^{-1} \nabla E_{\mathrm{med}}(X_r)$ (2b)
    $F \leftarrow F_{\mathrm{med}} + F_{\mathrm{slow}}$                      (2c)

    For $i = 1$ to $k_1$
        $X \leftarrow X + \frac{\Delta\tau}{2} V$                            (3a)
        $V \leftarrow V + \Delta\tau\,(F + F_{\mathrm{fast}}(X))$             (3b)
        $X \leftarrow X + \frac{\Delta\tau}{2} V$                            (3c)
    End

End
$V \leftarrow [V+\Delta\tau(F+F_{\mathrm{fast}}(X)+\mathbf{M}^{-1}R)]/(1+\gamma\Delta\tau)$   (3b*)
```

```
Impulse MTS
based on Velocity Verlet

$F_{\mathrm{slow}} \equiv -\mathbf{M}^{-1} \nabla E_{\mathrm{slow}}(X)$
$V \leftarrow V + \frac{\Delta t}{2} F_{\mathrm{slow}}(X)$

For $j = 0$ to $k_2 - 1$

    $F_{\mathrm{med}} \equiv -\mathbf{M}^{-1} \nabla E_{\mathrm{med}}(X)$
    $V \leftarrow V + \frac{\Delta t_m}{2} F_{\mathrm{med}}(X)$

    For $i = 0$ to $k_1 - 1$
        $V \leftarrow V + \frac{\Delta\tau}{2} F_{\mathrm{fast}}(X)$
        $X \leftarrow X + \Delta\tau\, V$
        $V \leftarrow V + \frac{\Delta\tau}{2} F_{\mathrm{fast}}(X)$
    End
    $V \leftarrow V + \frac{\Delta t_m}{2} F_{\mathrm{med}}(X)$
End
$V \leftarrow V + \frac{\Delta t}{2} F_{\mathrm{slow}}(X)$
```

Figure 4. Extrapolative versus Impulse MTS Approaches for Molecular Dynamics

For the LN scheme described in the text, the extrapolative MTS method is modified in three places: equation (1a) becomes $X_r^o = X$; after equation (3a), the random force *R* is evaluated; and equation (3b) is modified as shown in equation (3b*).

is still much less than the timescales of slower biomolecular motions (e.g., long-range electrostatics)!

The vulnerability of these impulse treatments to numerical stability became clear upon detailed analysis over the last five years [50–53]. This inherent *resonance*, or integrator-induced corruption of the system's dynamics, results from application of a force *impulse* (or pulse) at the onset and at the end of a sweep covering the long interval Δt: its strength is proportional to $\Delta t/\Delta\tau$, that is, *r* times larger in magnitude than those changes made to the position and velocity vectors in each inner timestep, $\Delta\tau$ (see Figures 3 and 4). Such resonance artifacts are integrator dependent [50–52].

Predictions for the occurrence of these resonances can be made based on analysis of harmonic models. Such analyses suggest that the most severe (third-order) resonance occurs when Δt is close to *half the fastest period* of the physical system [50]. In biomolecular systems, the first such resonance for Verlet-based MTS schemes occurs at about $\Delta t = 5$ fs, which is half the period of the fastest oscillation [50, 53]; this artifact can be delayed to the period, or 10 fs, when a stochastic formulation is used [11] (see Figure 3), but this is still much smaller than characteristic slow motions. Vivid illustrations of resonance artifacts in the dynamic simulations of proteins can be seen in Figure 3 for Newtonian and Langevin impulse splitting. In general, these instabilities in impulse-MTS treatments can be avoided only by reducing Δt, though implicit symplectic schemes can be devised to remove low-order resonances for model systems [52] (they are not practical, however, being too computationally demanding).

Fortunately, this understanding of resonance in *impulse* MTS methods suggested a combination of two simple ingredients that works together to alleviate resonances and allow larger timesteps in the LN scheme [37]: *extrapolative* force splitting (which by itself leads to energy drifts but is not as vulnerable to resonance artifacts as are impulse variants) and *stochastic dynamics* (which eliminate the drifts as well as dampen the mild resonances). The stochastic framework, in the form of Langevin dynamics [54], represents a departure from Hamiltonian dynamics, but stochastic simulations are appropriate for many thermodynamic and sampling questions, since the same equilibrium states are approached in theory. In practice, the bath-coupling parameter (damping constant γ; see Figure 2) can be set as small as possible to minimize the Hamiltonian perturbation while still ensuring numerical stability (e.g., γ around 10 ps^{-1} or less) [11, 55, 56]. This stochastic coupling was also adopted later by Skeel and coworkers, in the context of a mollified impulse method [57], but the impulse formulation restricts the outer timestep.

Validity and Applications

Performance of MTS schemes can be assessed by comparing energetic, geometric, and dynamic properties to STS simulations as well as experimental data where available. In the case of LN trajectories, the reference integrator is an extension of Verlet for Langevin dynamics [31] [see Figure 2, note below equation (7)].

The numerical stability and resonance alleviation were shown in trajectories for the proteins BPTI and lysozyme, a large water system [11, 55], solvated DNA [58], and a large DNA/protein system [56] (Yang et al., submitted), where STS results were well reproduced for outer timesteps of 50 fs or more. Results have shown that a good parameter choice for a 3-class LN scheme is $\Delta\tau = 0.5$ fs, $\Delta t_m = 1$ fs, and Δt up to 150 fs. If constrained dynamics (SHAKE [34]) for the light atom bonds are used, the inner timestep can be increased to 1 fs and the medium timestep to around 2 fs. The speedup factors depend on the reference system but can be an order of magnitude (factor of 10 or more) with respect to 0.5 fs inner timesteps [11] or around 5 for 1fs inner timesteps (when SHAKE is used). Relative mean energy differences between MTS and STS simulations are lower when SHAKE is applied [56].

To illustrate, LN's performance is shown in Figure 5 for three solvated biomolecular systems containing TATA box DNA (14 base pairs, 15320 atoms) [22], a polymerase β/DNA primer/template complex (43751 atoms) [56] (Yang et al., submitted), and TATA box DNA/TBP complex (41011 atoms) [D. Strahs, X. Qian, and T. S., unpublished data]. The "Manhattan plots," which report the differences in mean energy components as a function of the outer timestep Δt relative to STS Langevin simulations, show low relative errors in all energy components (below 3%) for outer timesteps up to $\Delta t = 120$ fs. In all cases, the inner and medium timesteps are fixed at 1 and 2 fs, respectively, and the computational speedup factor for $\Delta t = 120$ fs is four or more with respect to STS simulations at 1 fs.

The assignment of the Langevin parameter γ in the LN scheme ensures numerical stability on one hand and minimizes the perturbations to Hamiltonian dynamics on the other; we have used $\gamma = 10$ ps^{-1} or smaller in biomolecular simulations. To assess the effect of γ of dynamic properties, the protocol-sensitive spectral density functions computed from various trajectories can be analyzed [55] (Yang et al., submitted). As shown in Figure 5 for the pol β/DNA simulation, there is good agreement between the STS Langevin and LN-computed frequencies for the same γ (see also [11, 55, 56]; Yang et al., submitted). This agreement emphasizes the success of MTS integrators as long as the inner timestep is small. Furthermore, as γ is decreased, the characteristic frequencies generated by Langevin dynamics can more closely approximate Newtonian signals [11, 55, 56].

Detailed comparisons of the evolution of various geometric variables from each simulation (bottom dials in Figure 5) reflect the agreement between LN and the reference Langevin simulation as well. As expected, individual trajectories diverge, but the angular fluctuations are all in reasonable ranges.

The LN speedup factors are significant when large systems are investigated (4–7 compared to 1 fs single timesteps, see Figure 5). Added savings allow better estimates of minor-groove bending preferences in A-tract DNA [58] and a greater number of TATA variants (13) to be studied to assess sequence-dependent deformability and flexibility patterns relative to the wild-type TATA element [22]. The savings further allow sampling of many segments in the pathway of an enzyme opening motion—polymerase β complexed to primer/template DNA—with the open and closed crystallographic forms serving as experimental anchors (Yang et al., submitted). For the polymerase application, the accelerated sampling led to the identification of a key step in the pathway (Arg258 rotation, together with release of a catalytic magnesium ion), which is slow and may be rate limiting. This produced the intriguing hypothesis that the Arg258 rotation, rather than large subdomain movements per se, is a crucial aspect of pol β's DNA synthesis fidelity. A sequence of local motions involved in the pol β opening process was also suggested (Phe272 flip, pol β thumb movement, and the Arg258 rotation), along with the evolution of important hydrogen bonds and water molecules near the active site (Yang et al., submitted).

Perspective and Extensions

We have witnessed considerable progress over the past decade in the development of many rigorous and theoretically grounded MD integration algorithms, as well as analysis machinery for their interpretation. Problems like resonance or systematic drifts are now much better understood. Unfortunately, the transfer of these fundamental mathematical constructs to practical biomolecular simulations at large has been slow. This is not due so much to the complexity of the algorithms, but rather to the rapid pace of the production rate of MD trajectories caused by improvements in computer speed and parallelization, as well as the alternative sampling approaches that have been designed (e.g., Figure 1).

MTS methods, in particular, though in everyday use in several laboratories, have not yet superseded the STS Verlet scheme in popular packages like AMBER and CHARMM. This delayed transfer can be attributed to two specific areas that require further work: the optimal integration of MTS methods with PME methods for long-range electrostatics, and the application of MTS methods to various thermodynamic ensembles other than the microcanonical (NVE, or constant particle number, volume, and energy). Strategies have appeared in the literature, but the optimization of these designs in real simulations requires considerable effort.

The apparent difficulty in optimizing MTS/PME combinations is that the reciprocal term, which should isolate long-range slow forces, contains significant force contributions from near-field particle interactions [59]. This requires further mathematical transformations before adaptation to biomolecular MD. To clarify, recall that the summation for the electrostatic energy of a periodic system can be expressed by a lattice sum over all pair interactions and over all lattice vectors (excluding $i = j$ in the primary box). Such a sum is only conditionally convergent. Ewald's trick was to convert this sum into an expression involving a sum of an absolutely and rapidly convergent series in direct and reciprocal space. This transformation is accomplished by representing each point charge as a Gaussian charge density, producing an exponentially decaying function. This Gaussian transformation is counteracted by an analogous subtraction to leave the net result of an effective point charge. Thus, the electrostatic energy for a periodic system is expressed as a real-space (direct) term (energy due to point charges screened by oppositely charged Gaussians) that is short-range with a singularity at the origin, and an associated canceling term (periodic sum of Gaussians) that is smooth and long-range, summed in reciprocal space using smooth interpolation of Fourier series values [1]. In practice, the width of the Gaussian distribution is chosen to balance the work in the two terms: making the direct sum rapidly decreasing

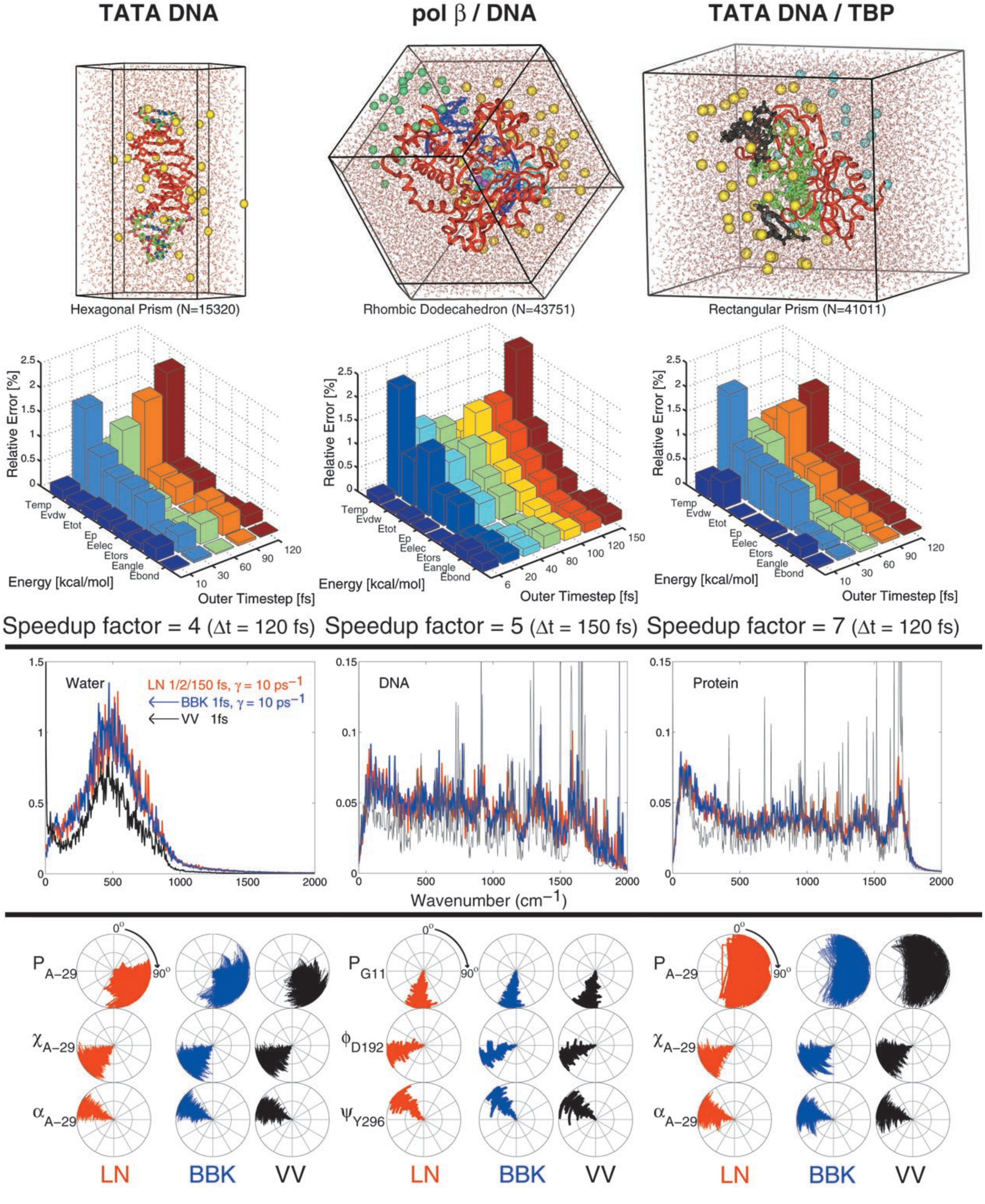


Figure 5. Performance of the Extrapolative Stochastic LN Scheme for Three Molecular Systems with Solvent and Salt

The three systems are a 14 bp TATA element, a pol β/DNA primer/template complex, and a DNA/TBP complex, each modeled in a different periodic domain that best fits the overall shape (top: hexagonal prism, rhombic dodecahedron, and rectangular box, respectively). The "Manhattan plots" show mean energy differences with respect to a single-timestep Langevin simulation as a function of LN outer timestep Δt (with $\Delta\tau = 1$ fs, $\Delta t_m = 2$ fs, $\gamma = 10$ ps^{-1}, and SHAKE applied to all light-atom bond stretches), as obtained from simulations for several picoseconds for each Δt. The spectral density functions are computed for the pol β/DNA system separately for water, DNA, and protein atoms by LN (1/2/150 fs, $\gamma = 10$ ps^{-1}), STS Langevin (1 fs, $\gamma = 10$ ps^{-1}), and velocity Verlet (1 fs, $\gamma = 0$); see computation protocol in [55]. Similarly, the evolution of selected geometric variables (sugar puckers, DNA backbone, and protein ϕ and ψ angles) for each simulation using LN STS Langevin, and velocity Verlet show good agreement over several picoseconds.

and well approximated for distances within a specified cutoff distance, and the reciprocal term converging rapidly. However, the rapid decay of the direct Ewald term implies the presence of fast components in the reciprocal term. This in turn means that the outer timestep associated with an MTS/PME code cannot be too large, as found in practice [60–62] (Batcho et al., submitted). Subsequently, truly efficient MTS/PME protocols likely need to split the direct (and hence modify the reciprocal) term so that the associated reciprocal term is truly long range.

Such ideas require alternative mathematical formulations, or a combination of distance and energy component criteria for defining the force classes. In addition, smoothing functions and possibly separate treatments for the van der Waals and Coulomb terms may be needed. Some ideas based on non-Gaussian core functions optimized to isolate the near- and far-field contributions in the direct and reciprocal terms, respectively, are presented in (Batcho and T.S., submitted), and another approach using an alternative splitting strategy for standard Gaussian-based PME formulations is currently being investigated [X. Qian and T. S., unpublished data].

These challenging extensions, as well as the efficient implementation of MTS schemes to constant temperature and pressure simulations [63], will undoubtedly be forthcoming in the near future and will help macromolecular models simulate a variety of physical environments, as appropriate, for less and less computer time. The ultimate bridging of the gap between experimental and simulation time frames will depend on the relative pace of improvements in each domain. To be sure, important challenges remain in each arena, and the sum of instrumentation and modeling will continue to be greater than each of its parts.

Acknowledgments

I thank Suse Broyde, Ron Elber, Hin Hark Gan, Jing Huang, John Kuriyan, Andy McCammon, Koneshan Siva, David Sept, Dan Strahs, Kaihsu Tai, Linjing Yang, Matthew Young, and Chung Wong for providing images for Figure 1, and Xiaoliang Qian, D. Strahs, and L. Yang for assembling the data used in Figures 3 and 5. Support by the National Science Foundation (ASC9157582, ASC9704681, BIR9318159), the National Institutes of Health (R01 GM55164), and a John Simon Guggenheim fellowship is gratefully acknowledged. T. Schlick is an investigator of the Howard Hughes Medical Institute.

References

1. Darden, T., Perera, L., Li, L., and Pedersen, L. (1999). New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations. Structure *7*, R55–R60.
2. Honig, B., and Nicholls, A. (1995). Classical electrostatics in biology and chemistry. Science *268*, 1144–1149.
3. Elber, R., Meller, J., and Olender, R. (1999). Stochastic path approach to compute atomically detailed trajectories: application to the folding of C peptide. J. Phys. Chem. B *103*, 899–911.
4. Board J.A., Jr., Kalé, L.V., Schulten, K., Skeel, R.D., and Schlick, T. (1994). Modeling biomolecules: larger scales, longer durations. IEEE Comp. Sci. Eng. *1*, 19–30.
5. Duan, Y., and Kollman, P.A. (1998). Pathways to a protein folding intermediate observed in a 1microsecond simulation in aqueous solution. Science *282*, 740–744.
6. Schlick, T., Barth, E., and Mandziuk, M. (1997). Biomolecular dynamics at long timesteps: bridging the timescale gap between simulation and experimentation. Annu. Rev. Biophys. Biomol. Struct. *26*, 179–220.
7. Schlick, T. (1999). Some failures and successes of long-timestep approaches for biomolecular simulations. In Computational Molecular Dynamics: Challenges, Methods, Ideas – Proceedings of the 2nd International Symposium on Algorithms for Macromolecular Modelling, Berlin, May 21–24, 1997, Volume 4 of Lecture Notes in Computational Science and Engineering, P. Deuflhard, J. Hermans, B. Leimkuhler, A.E. Mark, S. Reich, and R.D. Skeel, eds. (Berlin and New York: Springer Verlag), pp. 227–262.
8. Elber, R. (1996). Novel methods for molecular dynamics simulations. Curr. Opin. Struct. Biol. *6*, 232–235.
9. Doniach, S., and Eastman, P. (1999). Protein dynamics simulations from nanoseconds to microseconds*.* Curr. Opin. Struct. Biol. *9*, 157–163.
10. Daggett, V. (2000). Long timescale simulations. Curr. Opin. Struct. Biol. *10*, 160–164.
11. Barth, E., and Schlick, T. (1998). Overcoming stability limitations in biomolecular dynamics: I. combining force splitting via extrapolation with Langevin dynamics in LN. J. Chem. Phys. *109*, 1617–1632.
12. McCammon, J.A., and Harvey, S.C. (1997). Dynamics of Proteins and Nucleic Acids (Cambridge, MA: Cambridge University Press).
13. Brooks, C.L., III, Karplus, M., and Pettitt, B.M. (1998). Proteins: a Theoretical Perspective of Dynamics, Structure, and Thermodynamics, Volume LXXI of Advances in Chemical Physics (New York: John Wiley & Sons).
14. Gerstein, M., and Levitt, M. (1998). Simulating water and the molecules of life. Sci. Am. *279*, 101–105.
15. Ferrara, P., Apostolakis, J., and Caflisch, A. (2000). Targeted molecular dynamics simulations of protein unfolding. J. Phys. Chem. B. *104*, 4511–4518.
16. Madura, J.D., Davis, M.E., Gilson, M.K., Wade, R.C., Luty, B.A., and McCammon, J.A. (1994). Biological applications of electrostatic calculations and Brownian dynamics simulations. In Reviews in Computational Chemistry, Volume V, K.B. Lipkowitz and D.B. Boyd, eds., (New York: VCH Publishers), pp. 229–267.
17. Roux, B., and Simonson, T. (1999). Implicit solvent models. Biophys. Chem *78*, 1–20.
18. Bashford, D., and Case, D.A. (2000). Generalized Born models of macromolecular solvation effects*.* Annu. Rev. Phys. Chem. *51*, 129–152.
19. Schlick, T., Beard, D., Huang, J., Strahs, D., and Qian, X. (2000). Computational challenges in simulating large DNA over long times. IEEE Comp. Sci. Eng *2*, 38–51.
20. Jorgensen, W.L., and Tirado-Rives, J. (1996). Monte Carlo vs. molecular dynamics for conformational sampling. J. Phys. Chem. *100*, 14508–14513.
21. Schlick, T. (1992). Optimization methods in computational chemistry. In Reviews in Computational Chemistry, Volume III, K.B. Lipkowitz and D.B. Boyd, eds. (New York: VCH Publishers), pp. 1–71.
22. Qian, X., Strahs, D., and Schlick, T. (2001). Dynamic simulations of 13 TATA variants refine kinetic hypotheses of sequence/activity relationships. J. Mol. Biol., in press.
23. Young, M.A., Gonfloni, S., Superti-Furga, G., Roux, B., and Kuriyan, J. (2001). Dynamic coupling between SH2 and SH3 domains of C-Src and Hck underlies their inactivation by C-terminal tyrosine phosphorylation. Cell, in press.
24. Elcock, A.H., Gabdoulline, R.R., Wade, R.C., and McCammon, J.A. (1999). Computer simulation of protein-protein association kinetics: acetylcholinesterase-fasciculin. J. Mol. Biol. *291*, 149–162.
25. Gan, H.H., Tropsha, A., and Schlick, T. (2001). Lattice protein folding with two and four-body statistical potentials. Proteins *43*, 161–174.
26. Patel, D.J., et al., and Broyde, S. (1998). Nuclear magnetic resonance solution structures of covalent aromatic amine-DNA adducts and their magnetic relevance. Chem. Res. Toxic *11*, 391–407.
27. Huang, J., Schlick, T., and Vologodskii, A. (2001). Dynamics of site juxtaposition in supercoiled DNA. Proc. Natl. Acad. Sci. USA *98*, 968–973.

28. Verlet, L. (1967). Computer “experiments” on classical fluids: I. Thermodynamical properties of Lennard-Jones molecules. Phys. Rev. *159*, 98–103.
29. Hockney, R.W., and Eastwood, J.W. (1981). Computer Simulation Using Particles (New York:McGraw-Hill).
30. Swope, W.C., Andersen, H.C., Berens, P.H., and Wilson, K.R. (1982). A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: applications to small water clusters. J. Chem. Phys. *76*, 637–649.
31. Brünger, A., Brooks, C.L., III, and Karplus, M. (1982). Stochastic boundary conditions for molecular dynamics simulations of ST2 water. Chem. Phys. Lett. *105*, 495–500.
32. Störmer, C. (1907). Sur les trajectoires des corpuscules électrisés dans l’espace. Archives des Sciences Physiques et Naturelles *24*, 5–18, 113–158, 221–247. [This reference is the first in a three-part essay. The second part appeared in the same journal in 1911 (pp. 190, 277, 415, and 501), and the third part appeared in the 1912 volume of the journal (pp. 51–69)].
33. Sanz-Serna, J.M., and Calvo, M.P. (1994). Numerical Hamiltonian Problems (London: Chapman & Hall).
34. Ryckaert, J.P., Ciccotti, G., and Berendsen, H.J.C. (1977). Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of nalkanes. J. Comp. Phys. *23*, 327–341.
35. Andersen, H.C. (1983). Rattle: a “velocity” version of the SHAKE algorithm for molecular dynamics calculations. J. Comp. Phys. *52*, 24–34.
36. van Gunsteren, W.F. (1980). Constrained dynamics of flexible molecules. Mol. Phys. *40*, 1015–1019.
37. Barth, E., Mandziuk, M., and Schlick, T. (1997). A separating framework for increasing the timestep in molecular dynamics. In Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications, Volume III, W.F. van Gunsteren, P.K. Weiner, and A.J. Wilkinson, eds. (Leiden, The Netherlands: ESCOM) pp. 97–121.
38. Mazur, A.K., Dorofeev, V.E., and Abagyan, R.A. (1991). Derivation and testing of explicit equations of motion for polymers described by internal coordinates. J. Comp. Phys. *92*, 261–272.
39. Mathiowetz, A.M., Jain, A., Karasawa, N., and Goddard, W.A. (1994). Protein simulations using techniques suitable for very large systems: the cell multipole method for nonbonded interactions and the Newton-Euler inverse mass operator method for internal coordinate dynamics. Proteins *20*, 227–247.
40. Reich, S. (1996). Torsion dynamics of molecular systems. Phys. Rev. E *53*, 4176–4181.
41. Brünger, A.T., Adams, P.D., and Rice, L.M. (1999). Annealing in crystallography: a powerful optimization tool. Prog. Biophys. Mol. Biol. *72*, 135–155.
42. Streett, W.B., Tildesley, D.J., and Saville, G. (1978). Multiple timestep methods in molecular dynamics. Mol. Phys. *35*, 639–648.
43. Schlick, T., and Brandt, A. (1996). A multigrid tutorial with applications to molecular dynamics. IEEE Comp. Sci. Eng. *3*, 78–83.
44. Scully, J.L., and Hermans, J. (1993). Multiple timesteps: limits on the speedup of molecular dynamics simulations of aqueous systems. Mol. Sim. *11*, 67–77.
45. Grubmüller, H., Heller, H., Windemuth, A., and Schulten, K. (1991). Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions. Mol. Sim. *6*, 121–142.
46. Tuckerman, M.E., Berne, B.J., and Martyna, G.J. (1992). Reversible multiple time scale molecular dynamics. J. Chem. Phys. *97*, 1990–2001.
47. Biesiadecki, J.J., and Skeel, R.D. (1993). Dangers of multiple time-step methods. J. Comp. Phys. *109*, 318–328.
48. Zhou, R., and Berne, B.J. (1995). A new molecular dynamics method combining the reference system propagator algorithm with a fast multipole method for simulating proteins and other complex systems. J. Chem. Phys. *103*, 9444–9459.
49. Watanabe, M., and Karplus, M. (1995). Simulations of macromolecules by multiple timestep methods. J. Phys. Chem. *99*, 5680–5697.
50. Mandziuk, M., and Schlick, T. (1995). Resonance in the dynamics of chemical systems simulated by the implicit-midpoint scheme. Chem. Phys. Lett. *237*, 525–535.
51. Skeel, R.D., Zhang, G., and Schlick, T. (1997). A family of symplectic integrators: stability, accuracy, and molecular dynamics applications. SIAM J. Sci. Comp. *18*, 202–222.
52. Schlick, T., Mandziuk, M., Skeel, R.D., and Srinivas, K. (1998). Nonlinear resonance artifacts in molecular dynamics simulations. J. Comp. Phys. *139*, 1–29.
53. Barth, E., and Schlick, T. (1998). Extrapolation versus impulse in multiple-timestepping schemes: II. Linear analysis and applications to Newtonian and Langevin dynamics. J. Chem. Phys. *109*, 1632–1642.
54. Pastor, R.W. (1994). Techniques and applications of Langevin dynamics simulations. In The Molecular Dynamics of Liquid Crystals, G.R. Luckhurst and C.A. Veracini, eds. (Dordrecht, The Netherlands: Kluwer Academic) pp. 85–138.
55. Sandu, A., and Schlick, T. (1999). Masking resonance artifacts in force splitting methods for biomolecular simulations by extrapolative langevin dynamics. J. Comp. Phys. *151*, 74–113.
56. Schlick, T., and Yang, L. (2001). Long-timestep biomolecular dynamics simulations: LN performance on a polymerase β/DNA system. In Multiscale Computational Methods in Chemistry and Physics, volume 177, of NATO Science Series: Series III Computer and Systems Sciences, A. Brandt, J. Bernholc, and K. Binder, eds. (Amsterdam: IOS Press) pp. 293–305.
57. Izaguirre, J.A. (1999). Longer Time Steps for Molecular Dynamics. PhD thesis, University of Illinois at Urbana-Champaign, 1999. Also UIUC Technical Report UIUCDCSR992107. Available via http://www.cs.uiuc.edu/research/techreports.html.
58. Strahs, D., and Schlick, T. (2000). A-tract bending: insights into experimental structures by computational models. J. Mol. Biol. *301*, 643–666.
59. Heyes, D.M. (1981). Electrostatic potentials and fields in infinite point charge lattices. J. Chem. Phys. *74*, 1924–1929.
60. Figueirido, F., Levy, R.M., Zhou, R., and Berne, B.J. (1997). Large scale simulation of macromolecules in solution: Combining the periodic fast multipole method with multiple time step integrators. J. Chem. Phys. *106*, 9835−9849. Erratum published in. J. Chem. Phys. *107*, 7002.
61. Cheng, A., and Merz, K.M., Jr. (1999). Application of a multiple time step algorithm to biomolecular systems. J. Phys. Chem. B *103*, 5396–5405.
62. Kawata, M., and Mikami, M. (2000). Computationally efficient canonical molecular dynamics simulations by using a multiple time-step integrator algorithm combined with the particle mesh Ewald method and with the fast multipole method. J. Comp. Chem. *21*, 201–217.
63. Martyna, G.J., Tuckerman, M.E., Tobias, D.J., and Klein, M.L. (1996). Explicit reversible integrators for extended systems dynamics. Mol. Phys. *87*, 1117–1157.