31 December 2019

On The Importance of Accurate Algorithms for Reliable Molecular Dynamics Simulations

Berk Hess, David van der Spoel, Mark J. Abraham, Erik Lindahl[1]

1. KTH Royal Institute of Technology

Abstract

Molecular dynamics is expected to produce accurate results over a wide range of conditions and timescales. However, this is not always the case since the field has been too reluctant to abandon historically popular techniques known to introduce artefacts. Two recent papers have suggested there are reliability issues in the GROMACS code since it no longer uses a legacy twin-range algorithm. Here, we show there are order-of-magnitude differences in accuracy favoring the modern Trotter decomposition, and that a force field relying on the old algorithm will have errors parametrized into the force field. Similarly, the suggestions about incorrect virial calculations turn out to be explained by insufficient accuracy in the default SHAKE settings used for GROMOS, while the GROMACS default choices are accurate. This highlights the importance of being more critical to error cancellation in simulations in order for algorithms and parameters to both gradually converge to more perfect ones.

Keywords

molecular dynamics, force field, integration, Multiple Time Step, trotter

https://doi.org/10.26434/chemrxiv.11474583.v1

On The Importance of Accurate Algorithms for Reliable Molecular Dynamics Simulations

Berk Hess[1], David van der Spoel[2], Mark J. Abraham[1], Erik Lindahl[1,3]
[1]Department of Applied Physics, KTH Royal Institute of Technology, Stockholm, Sweden; [2]Department of Cell and Molecular Biology, Uppsala University, Uppsala, Sweden; [3]Department of Biochemistry and Biophysics, Stockholm University, Stockholm, Sweden

Abstract
Molecular dynamics is expected to produce accurate results over a wide range of conditions and timescales. However, this is not always the case since the field has been too reluctant to abandon historically popular techniques known to introduce artefacts. Two recent papers have suggested there are reliability issues in the GROMACS code since it no longer uses a legacy twin-range algorithm. Here, we show there are order-of-magnitude differences in accuracy favoring the modern Trotter decomposition, and that a force field relying on the old algorithm will have errors parametrized into the force field. Similarly, the suggestions about incorrect virial calculations turn out to be explained by insufficient accuracy in the default SHAKE settings used for GROMOS, while the GROMACS default choices are accurate. This highlights the importance of being more critical to error cancellation in simulations in order for algorithms and parameters to both gradually converge to more perfect ones.

Introduction
Molecular dynamics simulations have evolved over time and nowadays algorithms and force fields are expected to produce accurate results comparable to experiments over a wide range of conditions and timescales. However, as a community we have been too reluctant to abandon historically popular techniques that have later been shown to introduce artefacts, such as plain cutoffs or weak coupling algorithms[1,2]. One can understand the hesitation when substantial force field parametrization work might have been based on old algorithms with severe shortcomings – but at some point, we are forced to accept that certain parameters or setups are no longer adequate.

The GROMACS software[3] is part of this community, and the last few years there has been significantly increased interactions between major codes (including e.g. NAMD[4], AMBER[5], and CHARMM[6]) both on interoperability and what standards should be applied e.g. to accuracy. We are also fond of the legacy from the Berendsen laboratory where the first versions of GROMACS inherited parameters and algorithms from the GROMOS87[7] code and united-atom force field. With the advent of GROMOS96, those new potential forms and parameters were added to GROMACS as well. However, in parallel GROMACS early embraced a lot of development in the community at large, such as the particle-mesh Ewald algorithm for long-range electrostatics[8] (which we now even advocate for dispersion interactions[9,10]), and the code increasingly focused on all-atom force fields such as OPLS, AMBER and CHARMM.

The GROMOS family of force fields has up until the latest update[11] been developed with a twin-range (TR) cut-off. The GROMOS TR method computes short-range forces F_{SR} every step, but long-range force F_{LR} only every N^{th} step. F_{LR} is then added to F_{SR} *every* step to approximate a simulation with a longer cut-off. This setup can be better than a short cutoff, but still ignores interactions beyond the long cutoff which causes artefacts e.g. for lipid bilayers with large zwitterionic molecules[12]. A fundamental challenge with the legacy TR

setup is that the application of a force from an earlier step breaks time reversibility and means one is no longer integrating any well-defined Hamiltonian. Already in the 1990s Tuckerman et al. showed that for physically correct integration one needs to use a different kind of algorithm based on Trotter decomposition[13]. This is not merely theoretical hair-splitting for small systems, but the symplectic property of the Trotter decomposition means it exactly integrates a slightly perturbed Hamiltonian, which makes properties largely insensitive to the exact time step used rather than tuning parameters to fit a specific arbitrary time-step. This is why GROMACS adopted this algorithm a decade ago[3] instead of the legacy twin-range algorithm. The Trotter-based algorithm[13] has been a great advantage for force fields parametrized from simulations with a well-defined Hamiltonian as it combines high performance with accurate integration.

Two recent papers from the GROMOS community have suggested there are reliability issues in the GROMACS software as a result of this change[14,15]. In short, both papers notice that the GROMOS force field that was developed in conjunction with the GROMOS software and legacy twin-range cutoffs does not produce the same results when used with Trotter decompositions as implemented in GROMACS a decade ago, and to the best of our knowledge other modern codes using multiple-time-step integration. While the observations are mostly correct, we disagree with the conclusions, and we argue this brings up some important points for the community at large.

Results

First, it is obviously important that any changes that can influence results are documented in release notes, which was the case for GROMACS 4.5 under the entry *"Symplectic Trotter Leap-Frog integrator for twin-range non-bonded interactions."*[16]. This was apparently missed by some users, as neither the junior nor senior authors of one of the papers appear to have consulted the notes before using the new version as judged from the somewhat annoyed paper title[14]. Still, in retrospect we could of course have emphasized this more; we prefer to help users avoid mistakes e.g. by using a new interface for new implementations while the old one is deprecated, and refuse old input (at the cost of some annoyance for other users) - this principle is now one that we generally try to apply.

Nevertheless, despite the oversight, the two papers[14,15] highlight the interdependence of force field and algorithms which is an exceptionally important second point. Although it has been shown that biomolecules simulated using the GROMOS force field are more stable when using particle-mesh Ewald summation for the electrostatic interactions rather than the legacy TR scheme[17], the only way to assess if a force field is still compatible with more accurate algorithms is by comparing a wide set of predictions and properties[18]. To this end, Gonçalves and co-workers embarked on a series of simulations of molecular liquids[15] and they find significant differences between the legacy TR scheme and the symplectic algorithms: simple liquid properties like density change depending on the algorithm. Since the legacy TR scheme is not time-reversible, that means there is an interdependence between the force field and the cut-off scheme used, which is something all simulation users should be aware and cautious of as it can, in principle, occur whenever an algorithm is changed.

On the other hand, we strongly disagree with the suggestion[14] that the two cut-off schemes are qualitatively equivalent algorithms with slightly different properties – it is rather our opinion that the legacy TR scheme is fundamentally flawed and should not be used in any modern simulation that claims to accurately integrate a physical potential. The easiest way to realize this is to simply look at the energy conservation. For a trivial SPC water system[19], the legacy

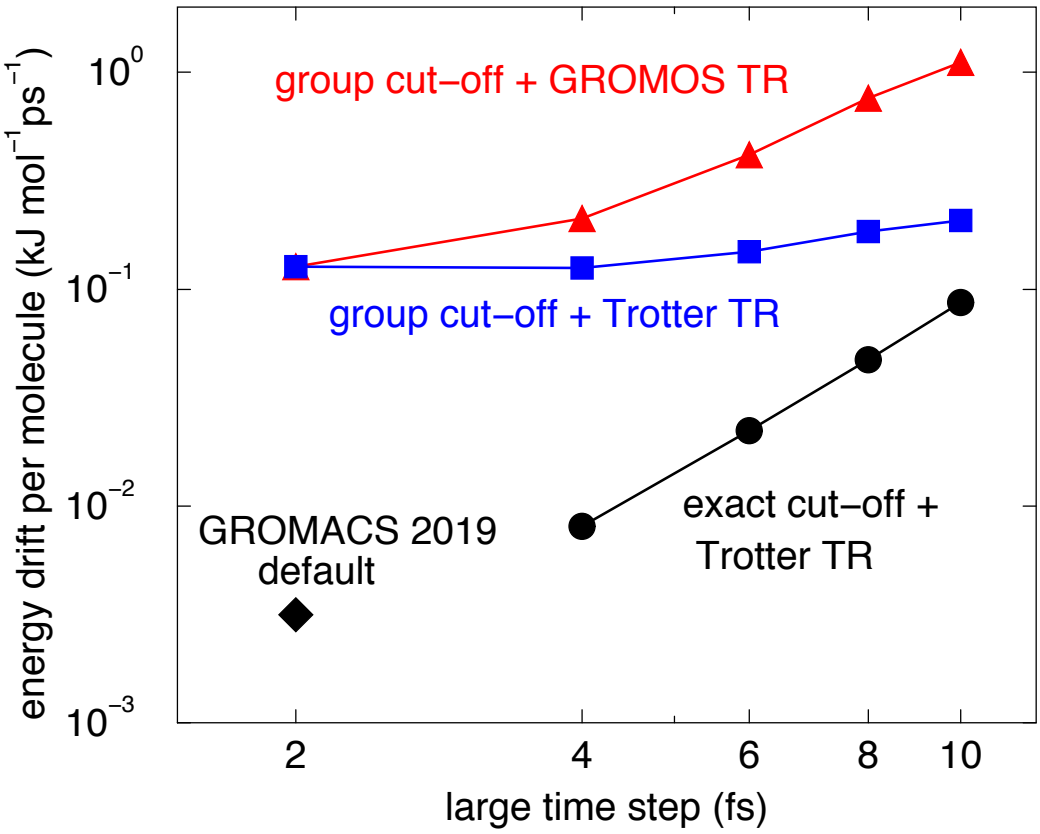

Figure 1: Energy drift per molecule for SPC water at 298.15 K and a density of 975 kg m^{-3} with reaction-field electrostatics with infinite ε_r, a long-range cut-off of 1.4 nm and a short time step of 2 fs. The comparison shows the legacy GROMOS twin-range cut-off scheme still used in GROMACS 4.0.7 (in 2009) and the new reversible Trotter twin-range scheme using GROMACS 5.1.2, the latter both for a group-based cut-off and an exact, atomic cut-off. The left-most points use a single 1.4 nm cut-off, the other points a 0.8/1.4 nm twin-range cut-off. Note how the group-based cut-off scheme causes alarmingly high (by modern standards) energy drift already without a twin-range scheme (here the neighbour list was updated every step, but updating it every 5 steps does not affect the drift). As the large (long-range) time step is increased to 10fs (the value used for GROMOS parametrization), the legacy TR (red) energy drift exceeds 1 kJ/mol/ps for each molecule. This has to be compensated with a thermostat, but such a strong coupling will have effects on kinetics. In contrast, while the Trotter TR scheme (blue) is still subject to the energy drift from the group-based cut-off per se, it is less sensitive to the actual size of the time step. For a single range exact atomic cut-off (black), there are no algorithmic errors and the energy only drifts due to rounding errors, which is the reason it increases as a power law with the time step, as expected (but from much lower values). For reference we also show the energy drift with default (exact cut-off) settings for GROMACS version 2019, which is 2.5 orders of magnitude lower than the legacy GROMOS TR scheme.

TR scheme with a 0.8/1.4 nm twin-range cut-off leads to an energy drift per molecule of 1.1 kJ mol^{-1} ps^{-1}, which is 0.45 k_BT per picosecond! (Fig. 1). This gigantic deviation from the correct Hamiltonian leads to an energy drift that is removed in the GROMOS scheme by very tight coupling using a Berendsen thermostat (another algorithm that, in our opinion, is no longer adequate[1]) with a coupling time of 0.1 ps. This high damping by the thermostat artificially stabilizes the integration scheme that uses an overly large time step of 10 fs for the long-range interactions. To confirm this, one can simply test it by instead combining the charge-group cut-off scheme with a single-range cut-off, which reduces the energy drift by an order of magnitude. It is technically possible to parametrize a force field with the legacy TR scheme, and some properties (in particular parametrization targets) will still be correct, but the specific (high) amount of leaky integration error that comes from the particular choice of time steps, cut-offs and integrator has then been parametrized into the force field. Apart from not integrating any well-defined Hamiltonian, as soon as a time step is changed (even making it shorter to aim for higher accuracy), the balance and energy drift will change and make the

results deviate from the intended target. We emphasize that many of these algorithms were state-of-the-art when they were developed, they contributed to many important scientific results, and for a long time many molecular dynamics codes (including ours) relied on this tweaking or cancellation of errors.

However, science makes progress by embracing new discoveries and learning from past shortcomings, even when a lot of work was invested in the latter. If one uses a symplectic Trotter integration scheme with an exact cut-off, as has been the default in GROMACS since the 5.0 release in 2014, the energy drift error is reduced by more than two orders of magnitude compared to the legacy TR scheme (Fig. 1). This in turn means the thermostat coupling will also be two orders of magnitude less tight, which has paramount impact on correctly integrating the kinetics without perturbing the system. This is important for determining fluctuation properties that may serve to evaluate force field accuracy[10,18], and as recently shown in the SAMPL6 challenge, free energy calculations are critically dependent on physically accurate integration[20]. As a bonus, the CPU performance of GROMACS 2019 is equal to the legacy TR scheme in GROMACS 4.0.7 for water and up to twice as fast for other compounds. In other words, using the correct physics-based integration algorithms not only improves accuracy, but also enables higher performance, which is another reason these algorithms have been adopted in most major MD codes today. As a side note, since the PME algorithm for both Coulomb- and Lennard-Jones interactions has now become so ubiquitous that it is a *de facto* standard for MD simulations[8,9] the TR multiple-time-stepping scheme (even the one based on correct Trotter decomposition) was completely removed in version 2016 of GROMACS, as the plain cutoffs themselves have severe physical shortcomings and have been superseded by PME.

Furthermore, the authors[15] claimed to have found a potential issue with the implementation of the LINCS constraint algorithm[21] in GROMACS 4.0.7. Although it is of course within the scientific freedom of the authors to report suspected errors in a paper, we think it would have been more productive to first discuss an observation with the developers to confirm, potentially adress it, and make sure other users are informed as well.

Since we feel obliged to do our due diligence, we have carefully checked the LINCS and SHAKE[22] algorithms in the 10 year old 4.0.7 version of GROMACS and found no issues with the virial calculation. Interestingly, though: when running the ethanol system with the settings used by Gonçalves et al.[15], we reproduced differences in the density between LINCS and SHAKE with the default settings used in GROMACS. We found that these differences are due to insufficient accuracy in SHAKE. With a relative tolerance of 10^{-4}, as used by Gonçalves et al.[15], which is the default value and the one used for parametrizing the GROMOS force field, the density is 776.5 ±0.1 kg m^{-3}. If one instead sets a tighter SHAKE tolerance of 10^{-5} it reproduces the default LINCS settings, whith densities of 777.9 and 777.8 ±0.1 kg m^{-3}, respectively. This thus appears to be another algorithm (or rather, settings) error that has been parametrized into the GROMOS force field. If the authors had asked us or submitted a bug report we would gladly have helped them investigate this to avoid making an incorrect claim in their paper. GROMACS has used LINCS by default for more than two decades, so simulations using other force fields are not affected (and to be fair, when constraining only bonds with hydrogens, a tolerance of 10^{-4} does not affect the density within the statistical accuracy of 0.1 kg m^{-3}, so this might mostly be a theoretical issue).

We think the GROMOS force field development should consider adopting modern algorithms and tighter tolerances for the settings in future releases of the force field. A comparison of

twin-range with single cut-off for the GROMOS 54A7 force field shows that the difference for amino acids seems to be small. It is not too far-fetched to expect that a force field optimized in conjunction with correct physical algorithms would improve accuracy and it is encouraging to see that some of the GROMOS developers indeed are adopting algorithms like particle-mesh Ewald for applications[23].

A word of caution to the community at large is that we have all been too hesitant about giving up old algorithms, and many of us that review papers have been too lenient towards accuracy issues. Molecular dynamics should not be a combination of unknown errors hidden by parameter tweaking, but a proper model of physics that gradually evolves towards more perfect algorithms and parameters. That does not mean old work or old papers were without merit, but we fail as a community if we do not recognize that some of the things we have all done historically are no longer sufficiently accurate by today's standards – and that, most likely, some of the things we do today will not be sufficient by tomorrow's standards.

Finally, it is both important and good practice to report software issues like suspected bugs straight to code developers. One of the advantages of free and open source software[24] that is also developed openly is the transparency of handling issues, including cases where the issue turns out not to be a bug. The GROMACS development team has handled issues openly for the last 13 years using Redmine (and previously Bugzilla), and we are very thankful for the testing, feedback and code improvements numerous users have contributed there.

Acknowledgments
This work was supported by the Swedish Research Council and the BioExcel CoE (H2020-INFRAEDI-02-2018-823830). Computational resources were provided by the Swedish National Infrastructure for Computing.

References

(1) Morishita, T. Fluctuation Formulas in Molecular-Dynamics Simulations with the Weak Coupling Heat Bath. *J. Chem. Phys.* **2000**, *113*, 2976–2982.
(2) Rosta, E.; Buchete, N.-V.; Hummer, G. Thermostat Artifacts in Replica Exchange Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2009**, *5* (5), 1393–1399. https://doi.org/10.1021/ct800557h.
(3) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; Van Der Spoel, D.; et al. GROMACS 4.5: A High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29* (7), 845–854. https://doi.org/10.1093/bioinformatics/btt055.
(4) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comp. Chem.* **2005**, *26*, 1781–1802.
(5) Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; T.E. Cheatham, I.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; et al. *AMBER 2018,*; University of California, San Francisco, 2018.
(6) Brooks, B. R.; Brooks III, C. L.; Mackerell Jr., A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; et al. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30* (10), 1545–1614. https://doi.org/10.1002/jcc.21287.
(7) van Gunsteren, W. F.; Berendsen, H. J. C. GROMOS87 Manual. Biomos BV: Groningen, The Netherlands 1987.
(8) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103* (19), 8577–8593. https://doi.org/10.1063/1.470117.
(9) Wennberg, C. L.; Murtola, T.; Hess, B.; Lindahl, E. Lennard-Jones Lattice Summation in Bilayer Simulations Has Critical Effects on Surface Tension and Lipid Properties. *J. Chem. Theory Comput.* **2013**, *9* (8), 3527–3537. https://doi.org/10.1021/ct400140n.
(10) Fischer, N. M.; Van Maaren, P. J.; Ditz, J. C.; Yildirim, A.; Van Der Spoel, D. Properties of Organic Liquids When Simulated with Long-Range Lennard-Jones Interactions. *J. Chem. Theory Comput.* **2015**, *11* (7).

https://doi.org/10.1021/acs.jctc.5b00190.
(11) Horta, B. A. C.; Merz, P. T.; Fuchs, P. F. J.; Dolenc, J.; Riniker, S.; Hünenberger, P. H. A GROMOS-Compatible Force Field for Small Organic Molecules in the Condensed Phase: The 2016H66 Parameter Set. *J. Chem. Theory Comput.* **2016**, *12* (8), 3825–3850. https://doi.org/10.1021/acs.jctc.6b00187.
(12) Patra, M.; Karttunen, M.; Hyvönen, M. T.; Falck, E.; Lindqvist, P.; Vattulainen, I. Molecular Dynamics Simulations of Lipid Bilayers: Major Artifacts Due to Truncating Electrostatic Interactions. *Biophys. J.* **2003**, *84* (6), 3636–3645. https://doi.org/10.1016/S0006-3495(03)75094-2.
(13) Tuckerman, M.; Berne, B. J.; Martyna, G. J. Reversible Multiple Time Scale Molecular Dynamics. *J. Chem. Phys.* **1992**, *97*, 1990–2001.
(14) Reißer, S.; Poger, D.; Stroet, M.; Mark, A. E. Real Cost of Speed: The Effect of a Time-Saving Multiple-Time-Stepping Algorithm on the Accuracy of Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2017**, *13* (6), 2367–2372. https://doi.org/10.1021/acs.jctc.7b00178.
(15) Gonçalves, Y. M. H.; Senac, C.; Fuchs, P. F. J.; Hünenberger, P. H.; Horta, B. A. C. Influence of the Treatment of Nonbonded Interactions on the Thermodynamic and Transport Properties of Pure Liquids Calculated Using the 2016H66 Force Field. *J. Chem. Theory Comput.* **2019**, *15* (3), 1806–1826. https://doi.org/10.1021/acs.jctc.8b00425.
(16) *GROMACS 4.5 Release Notes*; 2010.
(17) Lange, O. F.; Van Der Spoel, D.; De Groot, B. L. Scrutinizing Molecular Mechanics Force Fields on the Submicrosecond Timescale with NMR Data. *Biophys. J.* **2010**, *99* (2). https://doi.org/10.1016/j.bpj.2010.04.062.
(18) Caleman, C.; Van Maaren, P. J.; Hong, M.; Hub, J. S.; Costa, L. T.; Van Der Spoel, D. Force Field Benchmark of Organic Liquids: Density, Enthalpy of Vaporization, Heat Capacities, Surface Tension, Isothermal Compressibility, Volumetric Expansion Coefficient, and Dielectric Constant. *J. Chem. Theory Comput.* **2012**, *8* (1). https://doi.org/10.1021/ct200731v.
(19) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. Interaction Models for Water in Relation to Protein Hydration. In *Intermolecular Forces*; Pullman, B., Ed.; D. Reidel Publishing, 1981; pp 331–342.
(20) Rizzi, A.; Jensen, T.; Slochower, D. R.; Aldeghi, M.; Gapsys, V.; Ntekoumes, D.; Bosisio, S.; Papadourakis, M.; Henriksen, N. M.; de Groot, B. L.; et al. The SAMPL6 SAMPLing Challenge: Assessing the Reliability and Efficiency of Binding Free Energy Calculations. *bioRxiv* **2019**, 795005. https://doi.org/10.1101/795005.
(21) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comp. Chem.* **1997**, *18*, 1463–1472.
(22) Berendsen, J.-P. R. G. C.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comp. Phys.* **1977**, *23*, 327–341.
(23) Ramos, M. C.; Horta, V. A. C.; Horta, B. A. C. Molecular Dynamics Simulations of PAMAM and PPI Dendrimers Using the GROMOS-Compatible 2016H66 Forcefield. *J. Chem. Inf. Model.* **2019**. https://doi.org/10.1021/acs.jcim.8b00911.
(24) Elofsson, A.; Hess, B.; Lindahl, E.; Onufriev, A.; van der Spoel, D.; Wallqvist, A. Ten Simple Rules on How to Create Open Access and Reproducible Molecular Simulations of Biological Systems. *PLOS Comput. Biol.* **2019**, *15* (1), e1006649.