Post-5.0 feature clean-up plan

05/13/2014 10:03 AM - Mark Abraham

Status: New
Priority: Normal
Assignee: Mark Abraham
Category: core library
Target version: 5.x

Description

General status summary:

- #1292 (removing features for 5.0) is complete as far as I could achieve it on my own. Thanks to those who did contribute. Main items that happened there are removal of PD, and clean up for planned removal of iterated code for MTTK constraints.

- The Verlet scheme will be close to feature-complete in 5.0. The main thing missing is support for user tables, which is one of the things Alfredo Metere is working on. Other minor issues are that the GB parallelization is probably broken for both group and cutoff (so until someone wants to put time into it again, we support GB after 5.0 only in all-vs-all mode), and adress is only implemented for the group scheme (which needs at least new kernels, which should be addressed after we produce a kernel-generator python script)

- We do want a release where the Verlet and group schemes are both as reliable as we can reasonably achieve, so that any issues of cross-validating protocols can be settled unambiguously. I see that as the main reason for having a 5.1 later this year. Many of the group-scheme protocols are not worth reproducing, however.

Suggested features to announce as deprecated in 5.0

- group scheme (grompp already announces this; remove after 5.1)
- twin-range multiple-time-step scheme (it's been broken since before 4.5 #1400, and my attempt to fix it lately has not clearly succeeded; remove after 5.0)
- Generalized Born other than all-vs-all (#1054; remove after 5.0)
- iteration scheme for MTTK with constraints (remove after 5.0)
- ensemble-averaged restraints (unless someone actually is using it in ~4.6 and has a setup showing that it still works - discussion at #1117; remove after 5.0)
- mdrun -testverlet (since Verlet will be the default and people should no longer be converting ancient .tpr files - discussion at #1424; remove after 5.0)
- heuristic updates for group scheme (complicates signalling, which currently hurts multi simulations and makes it hard to verify a solution for #692, and Berk hopes to implement a better version of this for the Verlet scheme; remove after 5.0)

I don't like breaking/removing things that work, but "keep it, it may once have worked" is not a reasonable barrier to progress for those of us working on things we need for the future. "Here, keep this working" can be reasonable! :-)

Subtasks:

- Feature # 1054: Improve implicit solvent parallelization
- Bug # 1117: ensemble-averaged distance restraints is probably broken
- Task # 1424: remove mdrun -testverlet option
- Feature # 1665: improve free energy non-bonded kernel performance

Related issues:

- Related to Gromacs - Task #1017: C++ Vector/Matrix classes
- Related to Gromacs - Feature #1193: planning for new trajectory file reading ...
- Related to Gromacs - Bug #860: Checkpoint not created upon reaching time give...
- Related to Gromacs - Task #1411: Future of thread_mpi

Associated revisions

Revision b2b75d45 - 09/10/2014 03:40 PM - David van der Spoel
Remove .tpa, .tpb, .tpx, .trj files. Part of #1500.

Change-Id: Ia055cbf5311bf8780ca826c767207e5172d1b7e2
Related to the removal of the group scheme, the current code requires the use of group for anything using shells/Drudes (fatal error in shellfc.c if not). My current Drude branch is close to being ready, and I was hoping to have it considered for 5.1, so I wanted to mention that it's very important for us to make sure the shell/Drude code is supported going forward. I don't know what all that will entail as far as the Verlet kernels, but I wanted to note it here.

Very happy to continue to support such functionality, and happy to consider the new stuff for 5.1, but we will need to have a plan to port things to the Verlet scheme, and execute it later this year.

Other proposed things to deprecate and remove:

sd2 integrator (temperature is known to be too high; noticed by various people at various times, connected with #1498 though no report is there. The new sd integrator at https://gerrit.gromacs.org/#/c/3419 does better and is faster, and that patches announces the deprecation)

Should we deprecate some older compilers? E.g. ICC<13, gcc<4.6, clang<3.3. Of course users can still use them but we wouldn't fix compiler warnings or have pre-submit Jenkins tests. It is a bit annoying to have to fix compiler warnings for ancient compilers: e.g. http://jenkins.gromacs.org/job/Gromacs_Gerrit_master/5667/

Sounds good. Also MSVC2010 (at least?), and xlc<12.

Its good idea. Anyway most clusters admins and users installs their own recent compiler versions.

And perhaps some warnings when using an unsupported compiler version shipped on a so-called "enterprise" distributions.

Other proposed things to deprecate and remove:
In that case enterprise are so enterprise so for example ~30-40% of ki cluster users run gentoo prefix in their accounts or build some recent gcc version in $HOME

#14 - 05/26/2014 11:44 PM - Mark Abraham

More thoughts:

IMO, anything in contrib that looks like it might no longer work can go. If someone wants it, git knows about it, and if they do want it, either checking out the git version where they were committed, and/or doing some fix up is still going to be needed. I know I haven't been maintaining anything there during refactoring patches, though Teemu sometimes does.

Erik and I suggest that most of the analysis tools should move to contrib, to reflect their effective maintenance status. This will doubtless provoke some screaming, and need some wider discussion on gmx-developers, but I suggest

- tools should do one task
- that task must be unit tested
- tools must use the new analysis framework
- if the tool can't use the new analysis framework because the framework does not support enough functionality yet, then we can consider extending it (but the result doesn't have to be as sexy C++ as Teemu writes)
- thus, the tools that get ported will be the ones people want to use enough to put effort into maintaining - but because they only do one thing, and have a test, that maintenance will be easy-or-zero
- users will complain that it's harder to do their work, but that friction generates light and code, as well as heat

Making those conversions viable needs

- a trajectory-reading framework that implements something like the current functionality, behind a sane C++ API #1193
- a representation of rvec that can be used in C++ containers #1017 (various discussions elsewhere, but I suggest we pick something plausible for use in the analysis part of tools and just use it - Eigen looked good last time I considered the question. This will require things like the trajectory-reading code to read into rvec and then convert it, until someone cares enough to fix the situation, but that conversion is unlikely to bound the cost of the analysis. mdrun has separate needs and should be treated separately.)

#15 - 05/26/2014 11:44 PM - Mark Abraham

- Related to Task #1017: C++ Vector/Matrix classes added

#16 - 05/26/2014 11:45 PM - Mark Abraham

- Related to Feature #1193: planning for new trajectory file reading component added

#17 - 05/27/2014 07:34 AM - David van der Spoel

Just renaming gmxana to contrib does not have any effect. However, not compiling tools by default will mean we quickly lose users, since many users like gromacs for the analysis tools (as well as mdrun performance obviously). Unfortunately there are very few users that contribute code, right now. Most power users work towards their own scientific goals, including myself to a large extent. Looking in trajectoryanalysis/modules all but one of the tools have been coded by Teemu.

Unfortunately porting code to the new framework is a larger barrier than what one might think, it is pretty close to rewriting everything from scratch. Maybe we need to team up with some of our biggest users and get them to contribute more?
#18 - 05/27/2014 07:36 AM - David van der Spoel
By the way, the trajectory analysis framework already hides the actual trajectory reading from the tools, so this is taken care of - if the underlying code changes the tools are not affected. I'm implementing something similar for energy files with guidance from Teemu.

#19 - 06/03/2014 11:34 AM - Rossen Apostolov
Is anyone still using .g96 format? If it's obsolete lets remove it. I'll ask, just in case, on the mailing list before doing that.

#20 - 06/03/2014 11:38 AM - David van der Spoel
It is not entirely obsolete - I've used it recently as an intermediate when converting LAMMPS trajectories to gromacs, the reason being that g96 has higher precision than both pdb and gro and therefore it is easier to reproduce results from other packages. And pdb does not have velocities of course.

#21 - 06/03/2014 12:01 PM - Mark Abraham
I have used g96 on occasion, and some users certainly do. I don't recall if -f .gro -ndec <whatever> is just as good, though. Given that the implementation for .g96 is of similar quality to that of .gro, I'm happy to keep it. I was against .g87 and .xyz mostly on grounds of implementation quality.

#22 - 06/03/2014 03:51 PM - Rossen Apostolov
OK, lets keep it then.

Another format question - how about XMGR? Hasn't it been deprecated for a long time in favor of XMGRAcE? This is related to #783.

#23 - 06/03/2014 03:57 PM - Mark Abraham
I don't think xmgr is used much, but I don't think it's a problem either. Fixing #783 seems unrelated to removing xmgr.

#24 - 06/18/2014 02:26 PM - Mark Abraham
- Related to Bug #860: Checkpoint not created upon reaching time given in maxh added

#25 - 08/24/2014 01:01 PM - Mark Abraham
- Related to Task #1411: Future of thread_mpi added

#26 - 08/24/2014 01:04 PM - Mark Abraham
We should remove tpa, tpb and trj file formats. They've probably been unused for quite a few years, and it would surprise me if they worked correctly.
QM/MM support likely only works with group scheme, so it is also likely to disappear unless someone puts in some work.

More controversially, after thinking more about a comment from Erik the other day, I think we should dump thread-MPI sooner rather than later. There's multiple existing fallbacks for people needing to run on single nodes, and it does add complexity we don't absolutely need.

#27 - 09/09/2014 07:56 AM - Gerrit Code Review Bot
Gerrit received a related patchset '1' for Issue #1500.
Uploader: David van der Spoel (davidvanderspoel@gmail.com)
Change-Id: ia055cbf5311bf8780ca826c767207e5172d1b7e2
Gerrit URL: https://gerrit.gromacs.org/4017