GROMACS - Bug #2502

nonbonded interactions go missing with GPU when an empty domain goes non-empty

05/16/2018 09:59 PM - Roland Schulz

Status: Closed
Priority: High
Assignee: 
Category: core library
Target version: 2018.3
Affected version - extra info: 
Affected version: 2018

Description

Related issues:
Related to GROMACS - Bug #1990: LJ-PME unstable with OpenCL

Associated revisions
Revision 35d9ca98 - 08/22/2018 03:32 PM - Berk Hess

Fix missing interactions with GPU and DD

Non-local LJ and Coulomb interactions would not be computed on a rank after the non-local GPU pair-list was empty at some point in time; either at the start of the run or during a run.
The issue is that the pair-list was initialized conditionally on the size of the list in the device side data instead of the host side data.
Although this could have led to silent errors for small step numbers, most systems will likely crash for production runs.

Fixes #2502
Change-Id: lae1c5b70624b652d625520cadb647f862f296d5b

History
#1 - 06/26/2018 01:09 AM - Mark Abraham
Observed this again.

#2 - 06/26/2018 01:10 AM - Mark Abraham
- Related to Bug #1990: LJ-PME unstable with OpenCL added

#3 - 06/26/2018 01:16 AM - Mark Abraham
- Category set to core library

Perhaps needs the same kind of fix as at #1990? Is this perhaps the same kind of issue we have gotten before where our use of GPUs is not robust enough if DLB moves load fully away from a domain? (That should be tested with an integration test that perhaps constructs an empty domain and runs do_force_cutsVERLET on it.) Is it valid to call cudaMemcpyAsync with 0 buffer size? (CUDA runtime API is silent on that point) Or might we have left adat->f as nullptr in that case?

#4 - 06/26/2018 01:27 AM - Mark Abraham
Also seen in presubmit:

as

Error Message
Errors in checkpot.out (2 errors)
Stacktrace

checkpot.out:
comparing energy file ./reference_s.edr and ener.edr

There are 46 and 47 terms in the energy files

enm[13] (-- Conserved En.)
There are 11 terms to compare in the energy files

| LJ (SR)       | step 20: | 144.714, 142.008 |
|              | step 20: | -737.004, -739.709 |

and post-submit as on-demand:
http://jenkins.gromacs.org/job/Matrix_OnDemand/400/OPTIONS=gcc-5%20gpu%20ranks=4%20gpu_id=1%20cuda-8.0%20no-hwloc%20release-wit
h-assert%20host=bs_nix1204,label=bs_nix1204/testReport/(root)/complex/nbnxn_ljpme_LB/

Error Message
Errors in checkpot.out (2 errors)
Stacktrace

checkpot.out:
comparing energy file ./reference_s.edr and ener.edr

There are 46 and 47 terms in the energy files

enm[13] (-- Conserved En.)
There are 11 terms to compare in the energy files

| Angle         | step 20: | 205.527, 210.551 |
|              | step 20: | 2.52953, 3.34387 |

so one might guess that clearing of energy buffers is somehow inappropriate. But I imagine there are other possible causes.

#5 - 06/29/2018 10:56 AM - Mark Abraham

Berk also noticed

http://jenkins.gromacs.org/job/Matrix_PreSubmit_master/4546/OPTIONS=gcc-7%20gpu%20gpu_id=1%20cuda-9.2%20thread-mpi%20openmp%20c
make-3.6.1%20release-with-assert%20simd=avx2_256%20host=bs_nix1204,label=bs_nix1204/testReport/junit/(root)/complex/nbnxn_ljpme_LB_geo
metric/

as

checkpot.out:
comparing energy file ./reference_s.edr and ener.edr

There are 46 and 47 terms in the energy files

enm[13] (-- Conserved En.)
There are 11 terms to compare in the energy files

| Angle         | step 20: | 205.527, 210.551 |
|              | step 20: | 2.52953, 3.34387 |

Files read successfully

--------------------------------

checkforce.out:

v[ 0] ( 3.82190e-05 2.32143e-05 4.64865e-05) - ( 6.08868e-05 3.70635e-05 7.40042e-05)
| v[ 1] ( 5.81662e-06 8.28132e-06 -5.22912e-06) - (-2.88402e-05 -4.15874e-05 2.56450e-05)
| v[ 2] (-3.10117e-06 5.28137e-05 -1.75781e-05) - (-1.81094e-06 1.46093e-06 -1.00062e-06)
| v[ 7] ( 2.76249e-06 3.12851e-05 -5.39876e-06) - ( 3.02118e-06 2.82860e-05 -8.31177e-06)

02/22/2020
and


as

checkpot.out:
comparing energy file ./reference_s.edr and ener.edr
There are 46 and 47 terms in the energy files
enm[13] ( -- Conserved En.)
There are 11 terms to compare in the energy files
LJ (SR) step 20:  146.244, step 20:  140.421
Coulomb (SR) step 20:  -997.182, step 20: -1390.09
Potential step 20: -737.044, step 20: -1135.77
Files read successfully
--------------------------------
checkforce.out:

f[   0] (-1.94154e+02  3.57887e+02 -3.92766e+02) - (-1.91787e+02  3.59159e+02 -3.96034e+02)
f[   1] ( 1.59705e+01 -1.22633e+02 -2.74098e+02) - ( 1.82744e+01 -1.31923e+02 -2.91907e+02)
f[   2] ( 1.96284e+02  4.37623e+02  4.37623e+02) - ( 1.9550e+02  4.29007e+02  4.29007e+02)
f[   3] ( 4.56915e+02 -5.44059e+02 -6.46271e+02) - ( 4.55535e+02 -5.38697e+02 -6.37180e+02)
f[   4] ( 8.25692e+02  1.08274e+03  6.30820e+02) - ( 8.70024e+02  1.05985e+03  4.66785e+02)
f[   5] ( 4.57623e+02  8.71704e+02  4.34741e+02) - ( 4.54254e+02  8.41643e+02  4.69890e+02)
f[   6] ( 1.04015e+03  6.94798e+02  1.23552e+03) - ( 1.13498e+03  8.65864e+02  1.20288e+03)
f[   7] ( 2.74076e+02 -7.04603e+02 -4.74374e+02) - ( 2.57738e+02 -5.98203e+02 -5.94715e+02)
f[   8] ( 8.25692e+02  1.08274e+03  6.30820e+02) - ( 8.70024e+02  1.05985e+03  4.66785e+02)
f[   9] (-2.33992e+02  9.81907e+01  4.08109e+02) - ( 2.48566e+02  3.86293e+01  5.25807e+02)
f[  10] (-3.40616e+02  5.33086e+02  2.91028e+02) - (-3.78559e+02  5.34774e+02  3.03187e+02)

#6 - 06/29/2018 10:56 AM - Mark Abraham
- Subject changed from CUDA complex.nbnxn-ljpme-LB* post-submit tests unstable to CUDA complex.nbnxn-ljpme-LB* tests unstable

#7 - 06/29/2018 11:20 AM - Berk Hess
Note that some forces, which I assume are at step 0 as the velocities are near 0, are significantly off. Thus this is not an issue of energies being of at a certain step, but rather a (seemingly) correct energy evaluation, but incorrect forces at step 0 already. That must be some reduction or device to host communication/synchronization issue.

#8 - 06/29/2018 11:57 AM - Berk Hess
Ignore my previous comment.
There is only one set of force mismatches in the output, so the forces must match at step 0 and the mismatches are at step 20. That means that the forces are correct at step 0 and some or all forces are incorrect at one or more steps between 0 and 20.

#9 - 06/29/2018 01:42 PM - Berk Hess
I am wondering if this a actually related with LJ-PME-LB. The LJ-PME-LB kernels are identical to the LJ-PME-geom kernels, apart from the parameter lookup. That lookup is also used for the normal LJ-PME parameters without LJ-PME.
As only the PME-LB tests use the particular test system of one lipid and one water, I am wondering if this is not some other bug in the GPU code path unrelated to LJ-PME.

#10 - 07/02/2018 11:54 PM - Mark Abraham
The failing reports all use -gpu_id 1

#11 - 07/12/2018 06:47 PM - Mark Abraham
One theory I have is that this is similar to the empty domain issue we once had on fully loaded slaves - somehow a reduction of junk values from a no-work domain occurs

#12 - 07/12/2018 07:02 PM - Roland Schulz
You're theory sounds similar to what happened with OCL and required https://gerrit.gromacs.org/c/8058/6/src/gromacs/mdlib/sim_util.cpp . Maybe someone wants to look whether this change has an effect on CUDA.

#13 - 07/16/2018 07:00 PM - Szilárd Páll
I can reproduce the issue outside of jenkins, however it took quite many tries to get something that is more than a few % off. It is probably related to DD (can't reproduce at <3 ranks). Tested Roland's force clearing fix to #2404 and it does not eliminate the failure. Most often it's the nbnxn-ljpm-LB-geometric (rather than the nbnxn-ljpm-LB) test that fails, but this might not be too relevant. Could not reproduce however the nbnxn-vsites test failing that seems to happen in jenkins.

#14 - 07/30/2018 01:35 PM - Szilárd Páll
Note to self: we should check if it reproduces with -dlb no.

#15 - 08/08/2018 10:00 PM - Roland Schulz
swap is failing too on the same machine:

It might be that the FPE errors we got were related to this. Both are only on this configuration.

#16 - 08/08/2018 11:02 PM - Roland Schulz
And also nbnxn_pme_order6

#17 - 08/10/2018 07:31 AM - Berk Hess
- Status changed from New to In Progress
- Priority changed from Normal to High
- Target version set to 2018.3
- Affected version - extra info set to master
- Affected version changed from git master to 2018

This issue is more serious than we initially thought. All parallel GPU pair kernels are affected. After encountering zero non-local interactions, no non-local interactions will ever be evaluated on a rank.

#18 - 08/10/2018 09:07 AM - Gerrit Code Review Bot
Gerrit received a related patchset '1' for Issue #2502.
Uploader: Berk Hess (hess@kth.se)
Change-Id: gromacs~release-2018~Iae1c5b70624b652d625520cadb6471862f296d5b
Gerrit URL: https://gerrit.gromacs.org/8162

#19 - 08/10/2018 09:09 AM - Berk Hess
- Status changed from In Progress to Fix uploaded

Should we update the subject to reflect the more general issue?

#20 - 08/10/2018 12:16 PM - Peter Kasson
Slightly dumb & OT question--we noticed a situation yesterday (system a bit too complicated for a bug report, and I think there's an internal fix) where we have very different pressure coupling behavior when we run on CPUs vs. GPUs. Is it possible that a more subtle version of this could cause such behavior?

#21 - 08/10/2018 12:30 PM - Berk Hess
If that is a system with empty regions, I would certainly suggest to try if the fix resolves the problem. We have not noticed any issues in production and we have not had any reports from users, but we do not run a lot of systems with empty space.

02/22/2020
#22 - 08/10/2018 02:56 PM - Szilárd Páll
- Subject changed from CUDA complex.nlbxn-ijpm-LB* tests unstable to nonbonded interactions go missing with GPU when an empty domain goes non-empty
- Difficulty hard added
- Difficulty deleted (uncategorized)

I think this subject is more suitable.

#23 - 08/10/2018 03:42 PM - Gerrit Code Review Bot
Gerrit received a related patchset ‘1’ for Issue #2502.
Uploader: Szilárd Páll (pall.szilard@gmail.com)
Change-Id: gromacs~release-2018~I2e6875d1d6edf47e860c5b70cecc93e285f56815
Gerrit URL: https://gerrit.gromacs.org/8163

#24 - 08/22/2018 06:45 PM - Berk Hess
- Status changed from Fix uploaded to Resolved

Applied in changeset 35df9ca98c92c371aab02899e327797e4cbe457.

#25 - 08/23/2018 10:46 AM - Paul Bauer
- Status changed from Resolved to Closed