GROMACS - Task #1382

drop old CUDA support (3.2 and/or 4.0)

11/13/2013 04:47 PM - Szilárd Páll

Status:	Closed	
Priority:	Low	
Assignee:	Szilárd Páll	
Category:	mdrun	
Target version:	5.0	
Difficulty:	uncategorized	

Description

Dropping support for older CUDA version would not only allow code simplification, but also less testing. The versions in question are v3.2 (released Nov 2010) and v4.0 (May 2011).

Dropping v3.2 support would allow switching to page-locking memory using cudaHostRegister() instead of having to call cudaHotsMalloc() for the CPU-side memory that will be used in transfers to/from the GPU. Note that alignment and size requirements (page size dependent) apply.

Also dropping v4.0 support would allow removing the "legacy" kernels which are only used with the <=v4.1 CUDA versions.

While the former would make it impossible to build with CUDA 3.2, the drawback of the latter is only performance degradation. Also note that much of the Fermi hardware was launched in the 3.2-4.0 era and many are sticking to the old/original drivers like v270.x (older kernel dependency, stubbornness?) which means that they are limited to these older CUDA versions as newer CUDA requires newer drivers.

Related issues:

Related to GROMACS - Task #1456: remove the use of nbat->alloc/free pointers	**New**	**03/06/2014**

Associated revisions

Revision 2c0c359c - 12/03/2013 12:19 AM - Szilárd Páll

remove legacy CUDA non-bonded kernels

This commit drops the legacy set of kernels which were optimized for use with CUDA compilers 3.2 and 4.0 (previous to the switch to llvm backend in 4.1).

For now the only consequence is slight performance degradation with CUDA 3.2/4.0, the build system still requires CUDA >=3.2 as the kernels do build with the older CUDA compilers. Whether to require at least CUDA 4.1 will be decided later.

Refs #1382

Change-Id: I75d31b449e5b5e10f823408e23f35b9a7ac68bae

Revision 1bcf0ded - 03/11/2014 02:01 AM - Szilárd Páll

Bump required CUDA version to 4.0

- Added warning with CUDA 4.0;
- Fixed some typos in comments;
- Fixed the comment referring to GPU code generation with CUDA v5.0+.

Fixes #1382

Change-Id: Iee182ace765022fc6a179531d42e965594cff104

History

#1 - 11/14/2013 11:46 AM - Mark Abraham

Sounds good - resources are limited and we need to focus on the things that deliver value to us (i.e. performance on recent compute stacks).

#2 - 11/14/2013 12:29 PM - Szilárd Páll

+ we'll need a few more sets of kernels for shift/switch and hopefully energy groups for which we'd have to maintain a set of legacy kernels as well. Initially the work required to root out the legacy stuff may be rather high, but in the long term it will pay off.

After discussing with Berk, we decided that I'll drop a mail to gmx-users and gmx-dev to see whether there are any stuck with old CUDA.

However, note that I will prioritize features rather than this code maintenance - at least for 5.0.

#3 - 12/03/2013 04:09 PM - Szilárd Páll

- Status changed from New to In Progress

To complete this task we should require at least CUDA 4.0 or possibly 4.1.

#4 - 02/26/2014 01:11 AM - Szilárd Páll

- Target version changed from 5.x to 5.0

Bump the requirement to CUDA 4.0 and possibly remove the use of nbat->alloc/free pointers and instead use with CUDA the page-locking functionality in the CUDA RT API.

#5 - 02/28/2014 01:25 PM - Gerrit Code Review Bot

Gerrit received a related patchset '1' for Issue #1382.
Uploader: Szilárd Páll (pall.szilard@gmail.com)
Change-Id: Iee182ace765022fc6a179531d42e965594cff104
Gerrit URL: https://gerrit.gromacs.org/3203

#6 - 03/06/2014 09:05 PM - Szilárd Páll

- Related to Task #1456: remove the use of nbat->alloc/free pointers added

#7 - 03/11/2014 02:15 AM - Szilárd Páll

- Status changed from In Progress to Resolved

- % Done changed from 0 to 100

Applied in changeset 1bcf0ded90f494deb7ae70259d911985d8752077.

#8 - 04/03/2014 07:37 PM - Roland Schulz

- Status changed from Resolved to Closed