Update PME CUDA kernels to allow a different number of threads per atom in the gather and spread kernels.

10/30/2019 11:18 AM - Jonathan Vincent

Status: New
Priority: Normal
Assignee: 
Category: 
Target version: 
Difficulty: uncategorized

Description
Currently this is not supported because of the data layout for the data that is saved in the spread and reloaded in the gather has the data for each atom interleaved, i.e. data_1 atom_1, data_1 atom_2 ... data_1 atom_N, data_2 atom_1 etc, where there are N atoms per block. So the data layout is dependent on the number of atoms per block.

The data ordering most likely has to be changed to allow this.

Potentially the unit test code will also have to be updated as well as some unit tests just supply the spline and gridline data, which needs to be in the correct order

Associated revisions
Revision 22118220 - 10/30/2019 01:21 PM - Jonathan Vincent

Update PME CUDA spread/gather

Adds additional templated kernels to the CUDA spread and gather kernels. Allowing the use of 4 threads per atom instead of 16 and allowing the spline data to be recalculated in the spread instead of saved to global memory and reloaded.

The combinations mean we have 4 different kernels that can be called depending on which is most appropriate for the problem size and hardware (to be decided heuristically). By default existing method is used (16 threads per atom, saving and reloading of spline data).

Added an additional option to disable the preloading of charges and coordinates into shared memory, and instead each thread would deal with a single atom.

Removed the (currently disabled) PME_GPU_PARALLEL_SPLINE=1 code path.

Refs #2792 #3185 #3186 #3187 #3188

Change-Id: Ia48d8eb63e38d0d23eefd755dcc228ff9b68d3e6