

GROMACS - Task #2328

assess AVX2 128 vs 256 on AMD Zen and change defaults

12/04/2017 06:55 PM - Szilárd Páll

Status:	Closed																																												
Priority:	Normal																																												
Assignee:																																													
Category:	mdrun																																												
Target version:																																													
Difficulty:	uncategorized																																												
Description																																													
The performance difference between 128 and 256-bit builds on AMD Zen needs to be re-assessed and defaults changed in the light of recent more thorough testing.																																													
Initial assumption that 256-bit AVX2 will never be worth it turns out to not be correct because:																																													
<ul style="list-style-type: none">- Parts of the code (e.g. bonded, search) run a lot faster with 256-bit AVX, and many other kernels run with about the same performance- Nonbonded performance significantly improves with tabulated and/or 2xNN kernels																																													
Given the former GPU runs will often benefit from 256-bit AVX, e.g. octanol 55K with 1950X + Quadro P6000:																																													
	<table><thead><tr><th></th><th>1TPC</th><th>2TPC</th></tr></thead><tbody><tr><td>AVX2_128</td><td>81.69</td><td>84.859</td></tr><tr><td>AVX2_256</td><td>83.995</td><td>86.083</td></tr></tbody></table>		1TPC	2TPC	AVX2_128	81.69	84.859	AVX2_256	83.995	86.083																																			
	1TPC	2TPC																																											
AVX2_128	81.69	84.859																																											
AVX2_256	83.995	86.083																																											
Regarding CPU-only runs, some initial benchmarks indicate that tabulated Ewald kernels seem to be faster and possibly we might want to switch to 2xNN too with 256-bit SIMD:																																													
	<table><thead><tr><th></th><th colspan="2">water 48k</th><th colspan="2">octanol 55k</th></tr><tr><th></th><th>1TPC</th><th>2TPC</th><th>1TPC</th><th>2TPC</th></tr></thead><tbody><tr><td>AVX2_128 4xN ewTab</td><td>44.223</td><td>49.841</td><td>34.797</td><td>38.45</td></tr><tr><td>AVX2_128 4xN</td><td>44.165</td><td>47.124</td><td>34.246</td><td>36.141</td></tr><tr><td>AVX2_256 2xNN</td><td></td><td>43.45</td><td>46.926</td><td>34.963</td><td>36.762</td></tr><tr><td>AVX2_256 2xNN ewTab</td><td>42.702</td><td>48.085</td><td>34.031</td><td>37.71</td><td></td></tr><tr><td>AVX2_256 4xN ewTab</td><td>42.049</td><td>43.318</td><td>33.345</td><td>34.14</td><td></td></tr><tr><td>AVX2_256 4xN</td><td>42.961</td><td>43.148</td><td>33.641</td><td>33.811</td><td></td></tr></tbody></table>		water 48k		octanol 55k			1TPC	2TPC	1TPC	2TPC	AVX2_128 4xN ewTab	44.223	49.841	34.797	38.45	AVX2_128 4xN	44.165	47.124	34.246	36.141	AVX2_256 2xNN		43.45	46.926	34.963	36.762	AVX2_256 2xNN ewTab	42.702	48.085	34.031	37.71		AVX2_256 4xN ewTab	42.049	43.318	33.345	34.14		AVX2_256 4xN	42.961	43.148	33.641	33.811	
	water 48k		octanol 55k																																										
	1TPC	2TPC	1TPC	2TPC																																									
AVX2_128 4xN ewTab	44.223	49.841	34.797	38.45																																									
AVX2_128 4xN	44.165	47.124	34.246	36.141																																									
AVX2_256 2xNN		43.45	46.926	34.963	36.762																																								
AVX2_256 2xNN ewTab	42.702	48.085	34.031	37.71																																									
AVX2_256 4xN ewTab	42.049	43.318	33.345	34.14																																									
AVX2_256 4xN	42.961	43.148	33.641	33.811																																									
Related issues:																																													
Related to GROMACS - Bug #2327: AVX2_128 and AVX_128_FMA double precision PME...	Closed																																												

Associated revisions

Revision 1f860f69 - 12/07/2017 07:05 PM - Berk Hess

Choose faster nbxn SIMD kernels on AMD Zen

On AMD Zen tabulated Ewald kernels are always faster than analytical.
And with AVX2_256 2xNN kernels are faster than 4xN.
These faster choices are now made based on CpuInfo at run time.

Refs #2328

Change-Id: I146bc012910bc1f46ed14155651c3d2a7c1f91e5

Revision cacb3b41 - 12/10/2017 07:59 PM - Berk Hess

Remove SIMD warning for AMD Zen

After choosing nbxn 2xNN kernels and changing the to tabulated Ewald nonbonded kernels, AVX2_256 is only a few percent slower than AVX2_128 on AMD Zen and is faster with nonbondeds and PME on a GPU. So we should not warn the user when AVX2_256 is used.

Refs #2328

Change-Id: I67b66b0025c7e3c31943f3f02b80e97fb9764066

History

#1 - 12/04/2017 06:57 PM - Szilárd Páll

- Related to Bug #2327: AVX2_128 and AVX_128_FMA double precision PME gather regression added

#2 - 12/04/2017 11:15 PM - Gerrit Code Review Bot

Gerrit received a related DRAFT patchset '2' for Issue [#2328](#).

Uploader: Berk Hess (hess@kth.se)

Change-Id: gromacs~release-2018~I67b66b012910bc1f46ed14155651c3d2a7c1f91e5

Gerrit URL: <https://gerrit.gromacs.org/7283>

#3 - 12/05/2017 12:01 PM - Gerrit Code Review Bot

Gerrit received a related patchset '1' for Issue [#2328](#).

Uploader: Berk Hess (hess@kth.se)

Change-Id: gromacs~release-2018~I67b66b0025c7e3c31943f3f02b80e97fb9764066

Gerrit URL: <https://gerrit.gromacs.org/7289>

#4 - 12/05/2017 12:02 PM - Berk Hess

- Tracker changed from Bug to Task

- Status changed from New to Fix uploaded

- Affected version deleted (2018-beta1)

#5 - 12/05/2017 02:30 PM - Erik Lindahl

I don't expect this to be too system-dependent, but when making performance decisions we should include benchmarks for slightly more realistic systems with proteins or membranes.

#6 - 12/05/2017 11:53 PM - Szilárd Páll

Erik Lindahl wrote:

I don't expect this to be too system-dependent, but when making performance decisions we should include benchmarks for slightly more realistic systems with proteins or membranes.

Sure, I can do a broader set of tests, but note that these were intentionally picked because the octanol system contains about the same bonded work as a similarly-sized protein system would, but eliminates the impact of load imbalance with DD (which I also ran just did not include here). The water box was used because that exercises the special case where the nonbonded kernels can skip the LJ compute.

#7 - 12/11/2017 08:47 AM - Berk Hess

- Status changed from Fix uploaded to Resolved

This seems to be resolved for the moment. But (unfortunately) we will need to keep checking this when code changes or processors are updated.

#8 - 12/11/2017 11:46 AM - Erik Lindahl

- Status changed from Resolved to Closed