

GROMACS - Task #3360

investigate the future of DD dynamic load balancing with GPU offload

01/29/2020 11:23 AM - Szilárd Páll

| | | |
|---|--------|--------------------|
| Status: | New | |
| Priority: | Normal | |
| Assignee: | | |
| Category: | mdrun | |
| Target version: | | |
| Difficulty: | hard | |
| Description | | |
| <p>With most tasks offloaded there is increasingly less and less relevant CPU-side load measurement that the DD DLB can be based on. As CUDA still does not support GPU-side timing, we can't directly time the balanceable load in GPU runs with CUDA.</p> <ul style="list-style-type: none">• With most forces offloaded, but update on the CPU all we can rely on is the time the CPU spends waiting for forces from GPU which is not the most useful to balance on.• With update also offloaded, the CPU launch/schedule and GPU async execution completely go out of sync and therefore the CPU-side load measurements are completely meaningless. <p>We need to:</p> <ul style="list-style-type: none">• (on the short term) disable DLB for runs with GPU update and consider/investigate enabling cycle-based balancing• Investigate how can a new load balancing scheme be implemented (considering the limitation and features of GPU APIs) | | |
| Related issues: | | |
| Related to GROMACS - Feature #2891: PME/PP GPU communications | | In Progress |

History

#1 - 01/29/2020 11:23 AM - Szilárd Páll

- Related to Feature #2891: PME/PP GPU communications added