Assessing and tuning GROMACS performance on heterogeneous systems

Szilárd Páll
pszilard@kth.se

GROMACS on LUMI Workshop
January 25, 2024

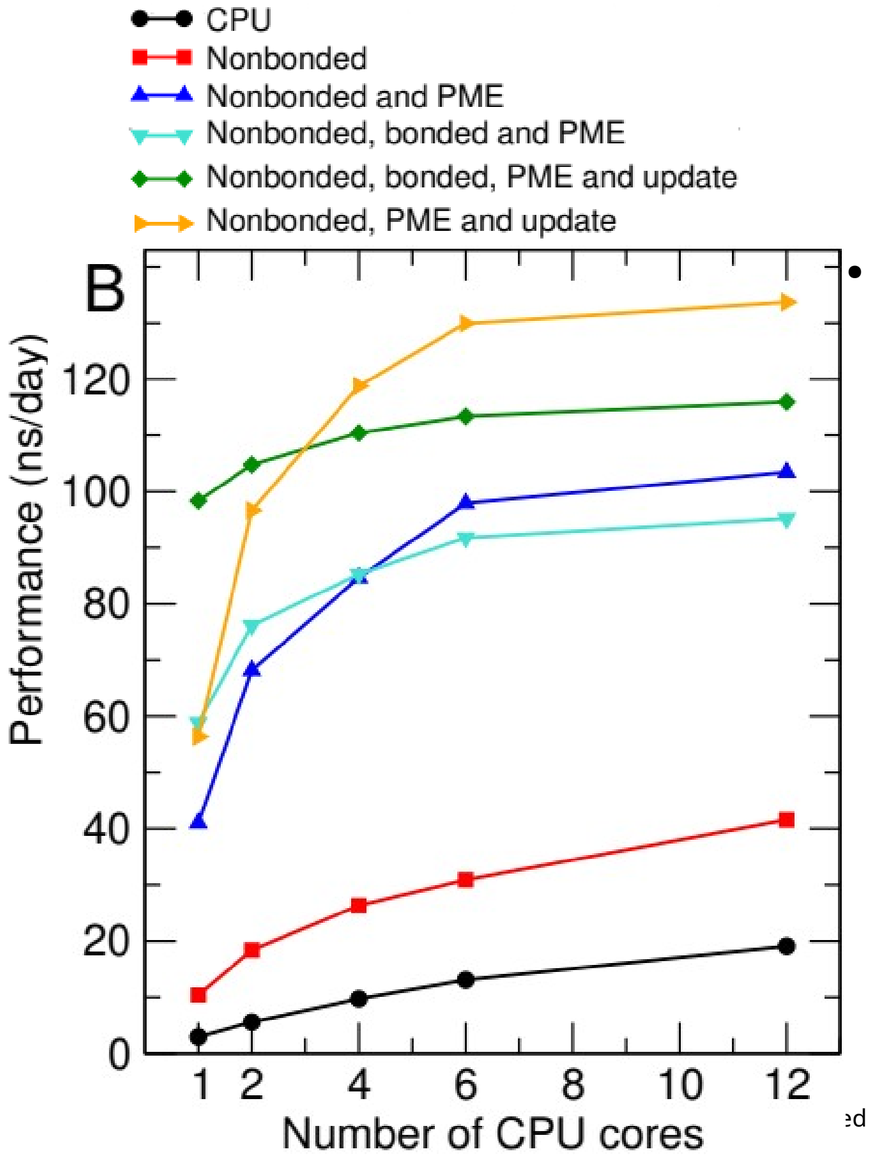
Why performance matters?

- **Improved time-to-solution**: get your results faster
  - wait less for your results
  - use your compute-hours effectively!
- **Energy efficiency**
  - faster time-to-solution on fixed hardware (num CPUs/GPU)

    ⇒ (typically) best energy-to-solution
  - faster time-to-solution on more hardware

    ~~⇒~~ not always best energy-to-solution

Why tuning performance matters: GPU parallelization modes



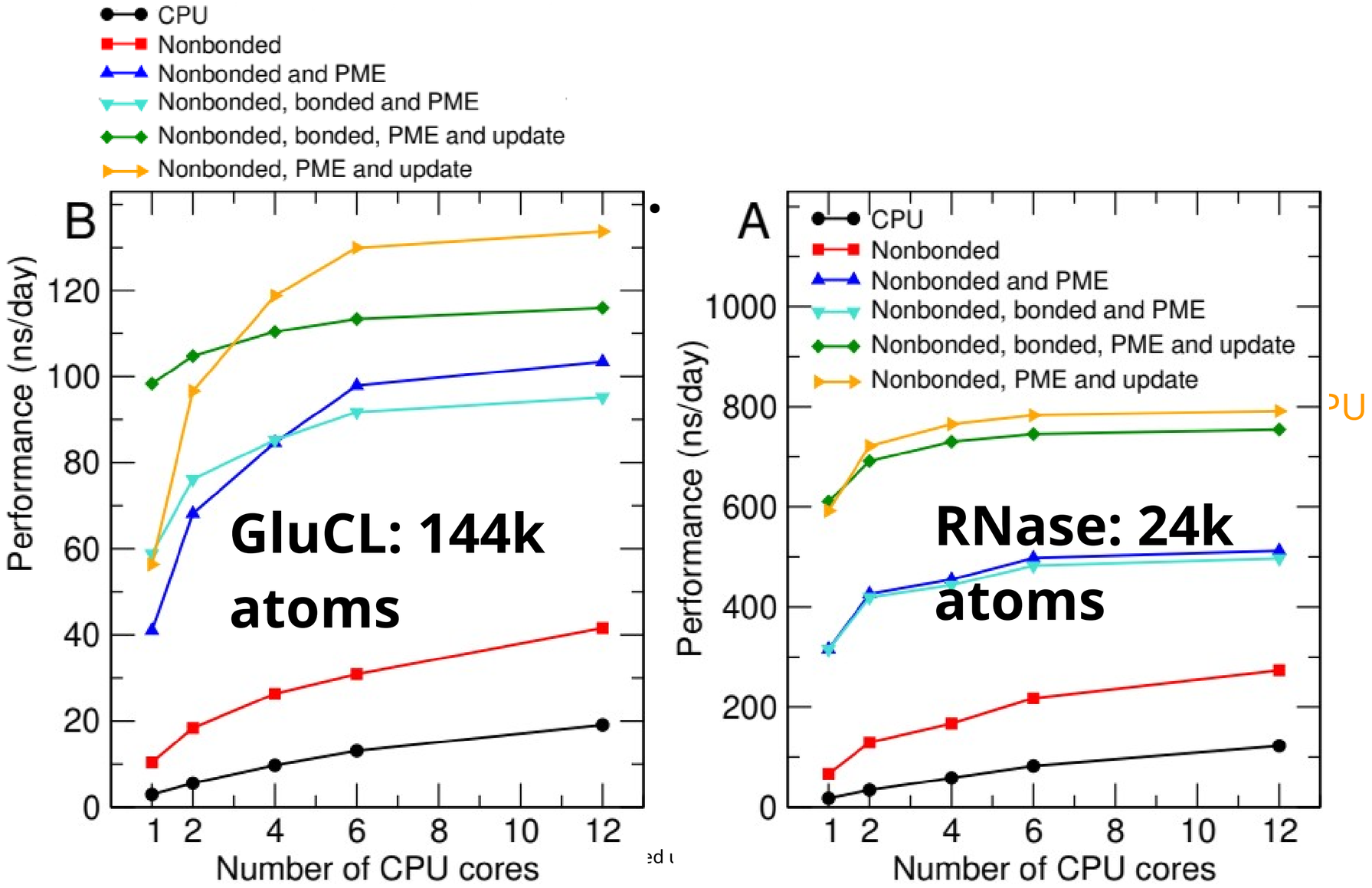
- Best performance:
  - with few cores/GPU: offload everything
  - from 3-4 cores/GPU: bondeds on CPU

Benchmark system: GluCL ion channel (144k atoms)

Hardware:

- AMD R3900X CPU
- NVIDIA 2080 SUPER GPU
- PCIe 3.0 interconnect (slow)

Why tuning performance matters: GPU parallelization modes



What influences performance?

- GROMACS version: use a recent release!

- Simulation setup:
  - system size
  - system settings (cutoff, long-range interactions, constraints, vsites, etc.)
    - check the documentation!
  - runtime options

- Compilers/libraries
  - some matter a lot: SYCL runtime, FFT library (GPU if offloaded)
  - other little (for simulation)

- Hardware: CPU/GPU/network

Reproducibility/repeatability

- Extensive reporting to fully document hardware, software & environment
  - Executable path + working dir
  - command line used
  - hardware used
  - MD simulation options (based on the MDP options)
  - env var-based features issue notes on their activation
  - algorihms used and their parameters
  - runtime/state

GROMACS version output

- Available through
  - gmx -v
  - every log file
- Allows identifying exact setup:
  - GROMACS version
  - build configuration
  - compilers/flags
  - libraries used

```
GROMACS version:    2023.3
Precision:          mixed
Memory model:       64 bit
MPI library:        MPI
OpenMP support:     enabled (GMX_OPENMP_MAX_THREADS = 128)
GPU support:        SYCL (hipSYCL)
NB cluster size:    8 (cluster-pair splitting off)
SIMD instructions:  AVX2_256
CPU FFT library:    commercial-fftw-3.3.10-sse2-avx-avx2-avx2_128
GPU FFT library:    VkFFT internal (1.2.26-b15cb0ca3e884bdb6c901a12d87aa8aadf7637d8) with HIP backend
Multi-GPU FFT:      none
RDTSCP usage:       enabled
TNG support:        enabled
Hwloc support:      disabled
Tracing support:    disabled
C compiler:         /appl/lumi/SW/LUMI-22.08/G/EB/rocm/5.3.3/llvm/bin/clang Clang 15.0.0
C compiler flags:   -mavx2 -mfma -Wno-missing-field-initializers -O3 -DNDEBUG
C++ compiler:       /appl/lumi/SW/LUMI-22.08/G/EB/rocm/5.3.3/llvm/bin/clang++ Clang 15.0.0
C++ compiler flags: -mavx2 -mfma -Wno-reserved-identifier -Wno-missing-field-initializers -Weverything -
Wno-c++98-compat -Wno-c++98-compat-pedantic -Wno-source-uses-openmp -Wno-c++17-extensions -Wno-documenta
tion-unknown-command -Wno-covered-switch-default -Wno-switch-enum -Wno-extra-semi-stmt -Wno-weak-vtables
 -Wno-shadow -Wno-padded -Wno-reserved-id-macro -Wno-double-promotion -Wno-exit-time-destructors -Wno-gl
obal-constructors -Wno-documentation -Wno-format-nonliteral -Wno-used-but-marked-unused -Wno-float-equal
 -Wno-cuda-compat -Wno-conditional-uninitialized -Wno-conversion -Wno-disabled-macro-expansion -Wno-unus
ed-macros -Wno-unused-parameter -Wno-unused-variable -Wno-newline-eof -Wno-old-style-cast -Wno-zero-as-n
ull-pointer-constant -Wno-unused-but-set-variable -Wno-sign-compare -Wno-unused-result -fopenmp=libomp -
O3 -DNDEBUG
BLAS library:       External - user-supplied
LAPACK library:     External - user-supplied
hipSYCL launcher:   /appl/local/csc/soft/chem/hipSYCL/0.9.4-cpeGNU-22.08/lib/cmake/hipSYCL/syclcc-launch
er
hipSYCL flags:      -Wno-unknown-cuda-version -Wno-unknown-attributes  --hipsycl-targets="hip:gfx90a"
hipSYCL GPU flags:  -ffast-math;-fgpu-inline-threshold=99999
hipSYCL targets:    hip:gfx90a
hipSYCL version:    hipSYCL 0.9.4-git
```

Hardware detection information

- Lists hardware details across all nodes in a simulation

  (if detection allows)

- On LUMI GPUs are “hidden” from MPI ranks (other than the one device the rank is assigned)

```
Running on 1 node with total 7 cores, 14 processing units, 1 compatible GPU
Hardware detected on host nid007958 (the node of MPI rank 0):
  CPU info:
    Vendor: AMD
    Brand:  AMD EPYC 7A53 64-Core Processor
    Family: 25   Model: 48   Stepping: 1
    Features: aes amd apic avx avx2 clfsh cmov cx8 cx16 f16c fma htt lahf misalignsse mmx msr
nonstop_tsc pcid pclmuldq pdpe1gb popcnt pse rdrnd rdtscp sha sse2 sse3 sse4a sse4.1 sse4.2 ss
se3 x2apic
  Hardware topology: Basic
    Packages, cores, and logical processors:
    [indices refer to OS logical processors]
      Package  0: [   1  65] [   2  66] [   3  67] [   4  68] [   5  69] [   6  70] [   7  71]
    CPU limit set by OS: -1   Recommended max number of threads: 14
  GPU info:
    Number of GPUs detected: 1
    #0: name: , architecture 9.0.10, vendor: AMD, device version: 1.2 hipSYCL 0.9.4-git, drive
r version 50322062, status: compatible
```

LUMI hardware detection: 1 node, 7 cores, 1 GPU

Hardware detection information

- Lists hardware details across all nodes in a simulation

  (if detection allows)

- On LUMI GPUs are “hidden” from MPI ranks other than the one the device is assigned to

```
Running on 4 nodes with total 224 cores, 448 processing units, 4 compatible GPUs
  Cores per node:           56
  Logical processing units per node:   112
  OS CPU Limit / recommended threads to start per node:   112
  Compatible GPUs per node:  1
  All nodes have identical type(s) of GPUs
Hardware detected on host nid007969 (the node of MPI rank 0):
  CPU info:
    Vendor: AMD
    Brand:  AMD EPYC 7A53 64-Core Processor
    Family: 25   Model: 48   Stepping: 1
    Features: aes amd apic avx avx2 clfsh cmov cx8 cx16 f16c fma htt lahf misalignsse mmx msr nonstop_tsc p
cid pclmuldq pdpe1gb popcnt pse rdrnd rdtscp sha sse2 sse3 sse4a sse4.1 sse4.2 ssse3 x2apic
  Hardware topology: Basic
    Packages, cores, and logical processors:
    [indices refer to OS logical processors]
      Package  0: [   1  65] [   2  66] [   3  67] [   4  68] [   5  69] [   6  70] [   7  71] [   9  73] [
  10  74] [  11  75] [  12  76] [  13  77] [  14  78] [  15  79] [  17  81] [  18  82] [  19  83] [  20  84
] [  21  85] [  22  86] [  23  87] [  25  89] [  26  90] [  27  91] [  28  92] [  29  93] [  30  94] [  31
 95] [  33  97] [  34  98] [  35  99] [  36 100] [  37 101] [  38 102] [  39 103] [  41 105] [  42 106] [
43 107] [  44 108] [  45 109] [  46 110] [  47 111] [  49 113] [  50 114] [  51 115] [  52 116] [  53 117]
[  54 118] [  55 119] [  57 121] [  58 122] [  59 123] [  60 124] [  61 125] [  62 126] [  63 127]
    CPU limit set by OS: -1   Recommended max number of threads: 112
  GPU info:
    Number of GPUs detected: 1
    #0: name: , architecture 9.0.10, vendor: AMD, device version: 1.2 hipSYCL 0.9.4-git, driver version 503
22062, status: compatible
```

LUMI hardweare detection: 4 node, 4x56 cores, 4x8 GPUs
(incorrectly reported as 4x1 because of ROCR_VISIBLE_DEVICES)

Hardware detection information

- Lists hardware details across all nodes in a simulation

  (if detection allows)

- On LUMI GPUs are “hidden” from MPI ranks other than the one the device is assigned to

```
Running on 1 node with total 128 cores, 256 processing units, 4 compatible GPUs
Hardware detected on host g1101.mahti.csc.fi:
  CPU info:
    Vendor: AMD
    Brand:  AMD EPYC 7H12 64-Core Processor
    Family: 23   Model: 49   Stepping: 0
    Features: aes amd apic avx avx2 clfsh cmov cx8 cx16 f16c fma htt lahf misalignsse mmx msr nonstop_tsc p
clmuldq pdpe1gb popcnt pse rdrnd rdtscp sha sse2 sse3 sse4a sse4.1 sse4.2 ssse3 x2apic
  Hardware topology: Basic
    Packages, cores, and logical processors:
    [indices refer to OS logical processors]
      Package  0: [   0 128] [   1 129] [   2 130] [   3 131] [   4 132] [   5 133] [   6 134] [   7 135] [
   8 136] [   9 137] [  10 138] [  11 139] [  12 140] [  13 141] [  14 142] [  15 143] [  16 144] [  17 145
] [  18 146] [  19 147] [  20 148] [  21 149] [  22 150] [  23 151] [  24 152] [  25 153] [  26 154] [  27
155] [  28 156] [  29 157] [  30 158] [  31 159] [  32 160] [  33 161] [  34 162] [  35 163] [  36 164] [
37 165] [  38 166] [  39 167] [  40 168] [  41 169] [  42 170] [  43 171] [  44 172] [  45 173] [  46 174]
[  47 175] [  48 176] [  49 177] [  50 178] [  51 179] [  52 180] [  53 181] [  54 182] [  55 183] [  56 18
4] [  57 185] [  58 186] [  59 187] [  60 188] [  61 189] [  62 190] [  63 191]
      Package  1: [  64 192] [  65 193] [  66 194] [  67 195] [  68 196] [  69 197] [  70 198] [  71 199] [
  72 200] [  73 201] [  74 202] [  75 203] [  76 204] [  77 205] [  78 206] [  79 207] [  80 208] [  81 209
] [  82 210] [  83 211] [  84 212] [  85 213] [  86 214] [  87 215] [  88 216] [  89 217] [  90 218] [  91
219] [  92 220] [  93 221] [  94 222] [  95 223] [  96 224] [  97 225] [  98 226] [  99 227] [ 100 228] [ 1
01 229] [ 102 230] [ 103 231] [ 104 232] [ 105 233] [ 106 234] [ 107 235] [ 108 236] [ 109 237] [ 110 238]
[ 111 239] [ 112 240] [ 113 241] [ 114 242] [ 115 243] [ 116 244] [ 117 245] [ 118 246] [ 119 247] [ 120 24
8] [ 121 249] [ 122 250] [ 123 251] [ 124 252] [ 125 253] [ 126 254] [ 127 255]
    CPU limit set by OS: -1   Recommended max number of threads: 256
  GPU info:
    Number of GPUs detected: 4
    #0: NVIDIA NVIDIA A100-SXM4-40GB, compute cap.: 8.0, ECC: yes, stat: compatible
    #1: NVIDIA NVIDIA A100-SXM4-40GB, compute cap.: 8.0, ECC: yes, stat: compatible
    #2: NVIDIA NVIDIA A100-SXM4-40GB, compute cap.: 8.0, ECC: yes, stat: compatible
```

CSC Mahti hardware detection: 1 node, 128 cores, 4 GPUs

Task assignment report

LUMI task assignment report

Note: this too is slighly misleading due to the “hidden” devices

```
On host nid007959 1 GPU selected for this run.
Mapping of GPU IDs to the 8 GPU tasks in the 8 ranks on this node:
  PP:0,PP:0,PP:0,PP:0,PP:0,PP:0,PP:0,PME:0
PP tasks will do (non-perturbed) short-ranged and most bonded interactions on the GPU
PP task will update and constrain coordinates on the GPU
PME tasks will do all aspects on the GPU
GPU direct communication will be used between MPI ranks.
Using 8 MPI processes
Using 7 OpenMP threads per MPI process
```

CSC Mahti task assignment report:

```
On host g1101.mahti.csc.fi 4 GPUs selected for this run.
Mapping of GPU IDs to the 4 GPU tasks in the 4 ranks on this node:
  PP:0,PP:1,PP:2,PME:3
PP tasks will do (non-perturbed) short-ranged and most bonded interactions on the GPU
PP task will update and constrain coordinates on the GPU
PME tasks will do all aspects on the GPU
GPU direct communication will be used between MPI ranks.
Using 4 MPI threads
Using 32 OpenMP threads per tMPI thread
```

- Reporting of task mapping
  - reported only for one node (of the first rank)
- Showing:
  - which tasks are offloaded
  - PP and PME task to GPU ID mapping

  Note that currently this can “break” due to ROCR_VISIBLE_DEVICES or equivalent

Domain decomposition report

```
Initializing Domain Decomposition on 8 ranks
Dynamic load balancing: auto
Using update groups, nr 389067, average size 2.7 atoms, max. radius 0.139 nm
Minimum cell size due to atom displacement: 2.438 nm
Initial maximum distances in bonded interactions:
    two-body bonded interactions: 0.442 nm, LJ-14, atoms 106625 106633
  multi-body bonded interactions: 0.442 nm, Proper Dih., atoms 106625 106633
Minimum cell size due to bonded interactions: 0.486 nm
Disabling dynamic load balancing; unsupported with GPU communication + update.
Using 1 separate PME ranks, as requested with -npme option
Optimizing the DD grid for 7 cells with a minimum initial size of 2.438 nm
The maximum allowed number of cells is: X 8 Y 8 Z 8
Domain decomposition grid 7 x 1 x 1, separate PME ranks 1
PME domain decomposition: 1 x 1 x 1
Interleaving PP and PME ranks
This rank does only particle-particle work.
Domain decomposition rank 0, coordinates 0 0 0

The initial number of communication pulses is: X 1
The initial domain decomposition cell size is: X 3.10 nm

The maximum allowed distance for atom groups involved in interactions is:
                 non-bonded interactions           2.436 nm
            two-body bonded interactions  (-rdd)   2.436 nm
          multi-body bonded interactions  (-rdd)   2.436 nm
```

Total rank count

DD cell size limits determines decomposition limits nstlist=400!

Dynamic load balancing not supported in GPU resident mode

Maximum decomposition setup possible (nstlist=400!)

Current PP/PME decomposition selected

Domain decomposition report: different nstlist

```
Initializing Domain Decomposition on 8 ranks
Dynamic load balancing: auto
Using update groups, nr 389067, average size 2.7 atoms, max. radius 0.139 nm
Minimum cell size due to atom displacement: 0.700 nm
Initial maximum distances in bonded interactions:
    two-body bonded interactions: 0.442 nm, LJ-14, atoms 106625 106633
  multi-body bonded interactions: 0.442 nm, Proper Dih., atoms 106625 106633
Minimum cell size due to bonded interactions: 0.486 nm
Disabling dynamic load balancing; unsupported with GPU communication + update.
Using 1 separate PME ranks, as requested with -npme option
Optimizing the DD grid for 7 cells with a minimum initial size of 0.700 nm
The maximum allowed number of cells is: X 30 Y 30 Z 30
Domain decomposition grid 7 x 1 x 1, separate PME ranks 1
PME domain decomposition: 1 x 1 x 1
Interleaving PP and PME ranks
This rank does only particle-particle work.
Domain decomposition rank 0, coordinates 0 0 0

The initial number of communication pulses is: X 1
The initial domain decomposition cell size is: X 3.10 nm

The maximum allowed distance for atom groups involved in interactions is:
                 non-bonded interactions           1.617 nm
            two-body bonded interactions  (-rdd)   1.617 nm
          multi-body bonded interactions  (-rdd)   1.617 nm
```

Total rank count

DD cell size limits determines decomposition limits nstlist=100!

Dynamic load balancing not supported in GPU resident mode

Maximum decomposition setup possible (nstlist=100!)

Current PP/PME decomposition selected

Pair interaction / Verlet algorithm setup

nstlist=100 (automatically chosen)

```
Using GPU 8x8 nonbonded short-range kernels

Using a dual 8x8 pair-list setup updated with dynamic, rolling pruning:
  outer list: updated every 100 steps, buffer 0.139 nm, rlist 1.339 nm
  inner list: updated every  12 steps, buffer 0.002 nm, rlist 1.202 nm
At tolerance 0.005 kJ/mol/ps per atom, equivalent classical 1x1 list would be:
  outer list: updated every 100 steps, buffer 0.292 nm, rlist 1.492 nm
  inner list: updated every  12 steps, buffer 0.051 nm, rlist 1.251 nm
```

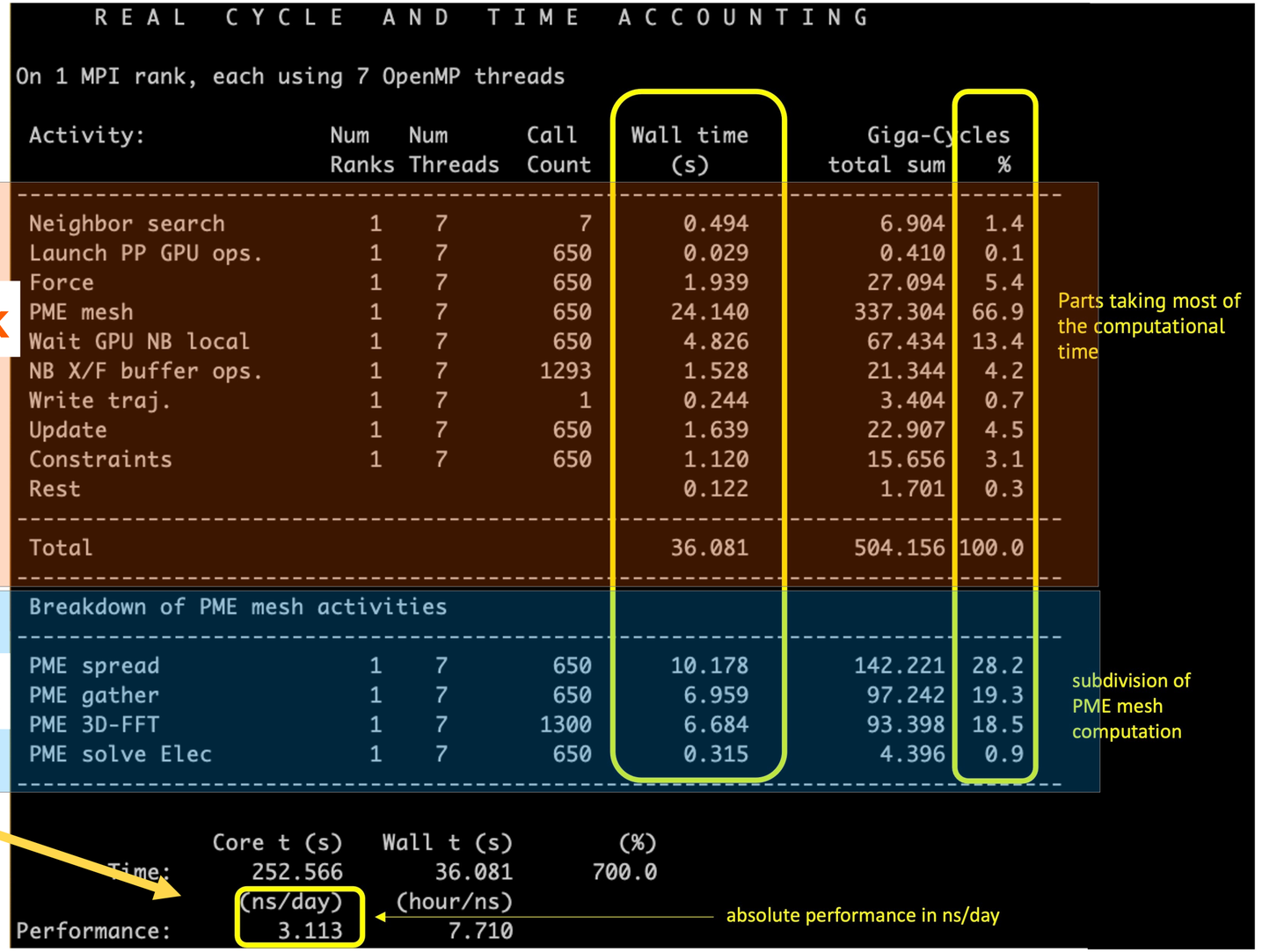
nstlist=400

```
Using GPU 8x8 nonbonded short-range kernels

Using a dual 8x8 pair-list setup updated with dynamic, rolling pruning:
  outer list: updated every 400 steps, buffer 0.958 nm, rlist 2.158 nm
  inner list: updated every  12 steps, buffer 0.002 nm, rlist 1.202 nm
At tolerance 0.005 kJ/mol/ps per atom, equivalent classical 1x1 list would be:
  outer list: updated every 400 steps, buffer 1.311 nm, rlist 2.511 nm
  inner list: updated every  12 steps, buffer 0.051 nm, rlist 1.251 nm
```

GROMACS performance table

- Displayed at the end of the run
- Timings of **CPU activities**
  - computation
  - communication
  - launch of GPU operations
  - waiting for data from GPU
- Final simulation performance



```
     R E A L   C Y C L E   A N D   T I M E   A C C O U N T I N G

On 1 MPI rank, each using 7 OpenMP threads

 Activity:                 Num    Num      Call    Wall time         Giga-Cycles
                           Ranks  Threads  Count       (s)         total sum    %
-----------------------------------------------------------------------------------
 Neighbor search             1    7            7       0.494          6.904   1.4
 Launch PP GPU ops.          1    7          650       0.029          0.410   0.1
 Force                       1    7          650       1.939         27.094   5.4
 PME mesh                    1    7          650      24.140        337.304  66.9
 Wait GPU NB local           1    7          650       4.826         67.434  13.4
 NB X/F buffer ops.          1    7         1293       1.528         21.344   4.2
 Write traj.                 1    7            1       0.244          3.404   0.7
 Update                      1    7          650       1.639         22.907   4.5
 Constraints                 1    7          650       1.120         15.656   3.1
 Rest                                            0.122          1.701   0.3
-----------------------------------------------------------------------------------
 Total                                          36.081        504.156 100.0
-----------------------------------------------------------------------------------
 Breakdown of PME mesh activities
-----------------------------------------------------------------------------------
 PME spread                  1    7          650      10.178        142.221  28.2
 PME gather                  1    7          650       6.959         97.242  19.3
 PME 3D-FFT                  1    7         1300       6.684         93.398  18.5
 PME solve Elec              1    7          650       0.315          4.396   0.9
-----------------------------------------------------------------------------------

               Core t (s)   Wall t (s)        (%)
       Time:      252.566       36.081      700.0
                 (ns/day)   (hour/ns)
Performance:        3.113       7.710
```

PP work

PME work

Parts taking most of the computational time

subdivision of PME mesh computation

absolute performance in ns/day

GROMACS performance table: multi-GPU run

- Displayed at the end of the run
- Timings of **CPU activities**
  - computation
  - communication
  - launch of GPU operations
  - waiting for data from GPU
- Final simulation performance

PP work

PME work

```
     R E A L   C Y C L E   A N D   T I M E   A C C O U N T I N G

On 7 MPI ranks doing PP, each using 7 OpenMP threads, and
on 1 MPI rank doing PME, using 7 OpenMP threads

 Activity:              Num   Num      Call    Wall time         Giga-Cycles
                        Ranks Threads  Count      (s)         total sum    %
-----------------------------------------------------------------------------
 Domain decomp.           7    7        256       3.108        303.821   8.7
 Send X to PME            7    7      12750       1.665        162.711   4.7
 Neighbor search          7    7        128       1.379        134.808   3.9
 Launch PP GPU ops.       7    7      50744       1.124        109.886   3.1
 Comm. coord.             7    7      12622      11.164       1091.134  31.2
 Force                    7    7      12750       0.016          1.558   0.0
 Wait + Comm. F           7    7      12750       5.084        496.927  14.2
 PME GPU mesh *           1    7      12750       7.001         97.747   2.8
 PME wait for PP *                                24.316        339.517   9.7
 Wait + Recv. PME F       7    7      12750       5.336        521.503  14.9
 Wait Bonded GPU          7    7         14       0.000          0.001   0.0
 Wait GPU NB nonloc.      7    7      12750       0.046          4.528   0.1
 Wait GPU NB local        7    7      12750       0.001          0.140   0.0
 Wait GPU state copy      7    7       2832       0.622         60.842   1.7
 NB X/F buffer ops.       7    7         28       0.003          0.329   0.0
 Write traj.              7    7          1       0.114         11.099   0.3
 Comm. energies           7    7       1275       0.817         79.893   2.3
 Rest                                             0.813         79.493   2.3
-----------------------------------------------------------------------------
 Total                                           31.294       3495.626 100.0
-----------------------------------------------------------------------------
(*) Note that with separate PME ranks, the walltime column actually sums to
    twice the total reported, but the cycle count total and % are correct.
-----------------------------------------------------------------------------
 Breakdown of PME mesh activities
-----------------------------------------------------------------------------
 Wait PME GPU gather      1    7      12750       0.042          0.590   0.0
 Launch PME GPU ops.      1    7     191250       0.622          8.691   0.2
 Wait PME Recv. PP X      1    7      89250       6.337         88.481   2.5
-----------------------------------------------------------------------------

               Core t (s)   Wall t (s)        (%)
       Time:     1751.068       31.294     5595.5
                 (ns/day)    (hour/ns)
Performance:       70.403        0.341
```

Assessing performance summary

- Acceptable vs reasonable performance / scaling
- Rough scaling guide
  - CPUs: hundreds of atoms / core
  - GPUs: tens of thousands of atoms / GPU
  - **Assuming**: “vanilla” MD setup and a high-performance interconnect
- Find reference data online and compare!
  - e.g. benchmark of similar simulation system on similar hardware
- Scaling: **check if it scales** do not just aassume
  - re-check with new input don’t just reuse settings
  - re-check if machine setup changes

Tuning performance: where to start?

- GROMACS version: use a recent release!

- Simulation setup:
  - system size
  - system settings (cutoff, long-range interactions, constraints, vsites, etc.)
    - check the documentation!
  - runtime options: reduce frequency of I/O and CPU-based algorithms (t/p coupling, comm motion removal)

- Compilers/libraries
  - some matter a lot: SYCL runtime, FFT library (GPU if offloaded)
  - other little (for simulation)

- Hardware: CPU/GPU/network

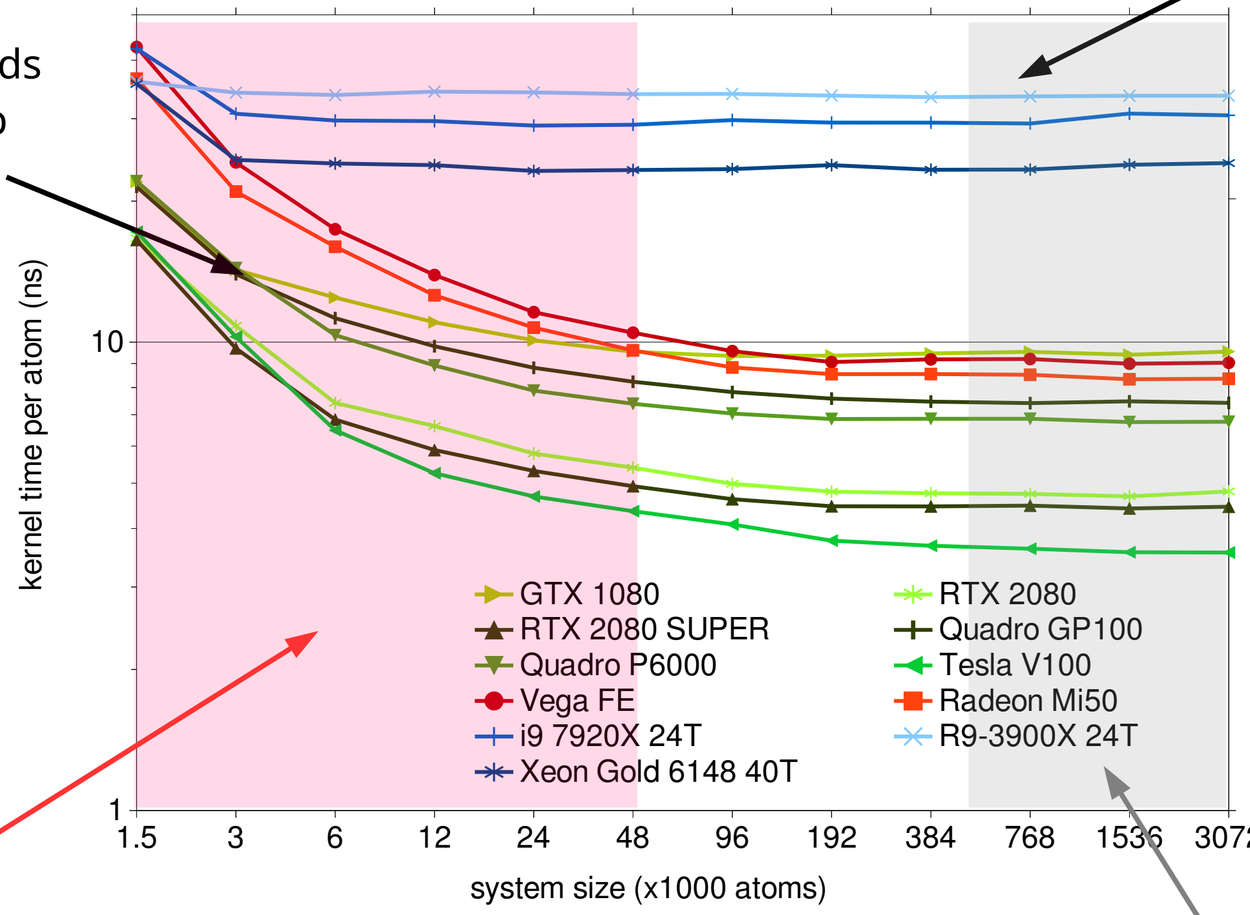
Tuning performance: what to do next?

- Check for features unsupported on GPUs/with GPU-resident mode
  - e.g. non “md” integrator, vsites,
- Make sure correct binding/affinities are used
  - suspicious sign: CPU tasks are taking unusually long
- Test offload modes:
  - prefer GPU-resident mode on modern hardware
- Use direct GPU comm
  - use a GPU-aware MPI
  - check for update groups (topology order issue: hydrogen directly after the heavy atom)
- Consider tunables:
  - nstlist
  - PP-PME balance, PP to PME GPU ratio
  - MPI ranks per GPU
  - OpenMP threads/rank

Anatomy of pair interaction kernel throughput

GPUs very sensitive to input size:
fixed overheads
kernel startup
SM load imabanlce

CPUs insensitive to input size to 100s atoms/core cache effects at large inputs



Strong scaling regime:
where most of our efforts go!

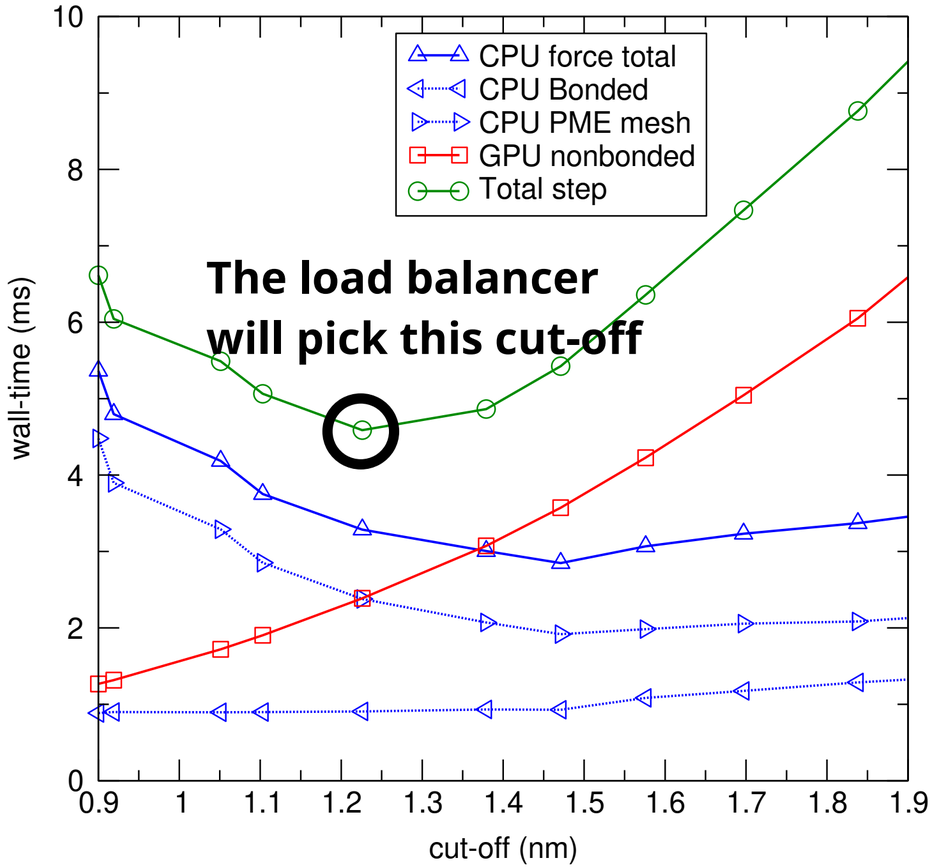
Benchmark “show-off” regime:

This is where the` **“free lunch” from new hardware** comes in full effect

PP-PME load balancing

- Task load balancing:
  - shift work from long- to short-range electrostatic
  - increase cutoff while decreasing grid spacing

- Used with:
  - MPMD : PP – PME ranks
  - non-bonded offload: CPU-GPU



GROMACS CPU-GPU balancing in practice

```
step   40: timed with pme grid 100 100 100, cutoff 0.900: 1671.1 M-cycles
step   80: timed with pme grid 84 84 84, cutoff 1.050: 1440.8 M-cycles
step  120: timed with pme grid 72 72 72, cutoff 1.225: 1879.7 M-cycles
step  160: timed with pme grid 96 96 96, cutoff 0.919: 1551.3 M-cycles
step  200: timed with pme grid 84 84 84, cutoff 1.050: 1440.7 M-cycles
step  240: timed with pme grid 80 80 80, cutoff 1.102: 1539.1 M-cycles
              optimal pme grid 84 84 84, cutoff 1.050
```

- Time consecutive cut-off settings, pick fastest
  - need to adjust PME grid ⇒ discrete steps
- Robust:
  - discard first timings
  - re-try if fluctuation is noticed
- Weaknesses:
  - (Computational tradeoff)
  - Static load balance: bias-prone by initial machine state, CPU/GPU clock ramp-up or thorttle

Further resources

- GROMACS documentation:
  - Getting good performance from mdrun
    https://manual.gromacs.org/documentation/current/user-guide/mdrun-performance.html#getting-good-performance-from-mdrun
  - Performance checklist:
    https://manual.gromacs.org/documentation/current/user-guide/mdrun-performance.html#performance-checklist
- S. Páll, et. al (2020). Heterogeneous Parallelization and Acceleration of Molecular Dynamics Simulations in GROMACS. J. Chem. Phys. 153, 134110 (2020);
  https://doi.org/10.1063/5.0018516
- Maximizing GROMACS Throughput with Multiple Simulations per GPU Using MPS and MIG
  https://developer.nvidia.com/blog/maximizing-gromacs-throughput-with-multiple-simulations-per-gpu-using-mps-and-mig
- Post your questions on the GROMACS users' forum: https://gromacs.bioexcel.eu

Performance vs search frequency

- nstlist free parameter
  - accuracy-based list buffering given the verlet-buffer-tolerance mdp parameter
- Dual pair list allows increasing nstlist to much larger values
  - automated Verlet buffer is needed

  (does not work with verlet-buffer-tolerance=-1