Efficient (GROMACS+)CP2K compute resource usage for QM/MM simulation of biomolecular systems

Arno Proeme Holly Judge

EPCC

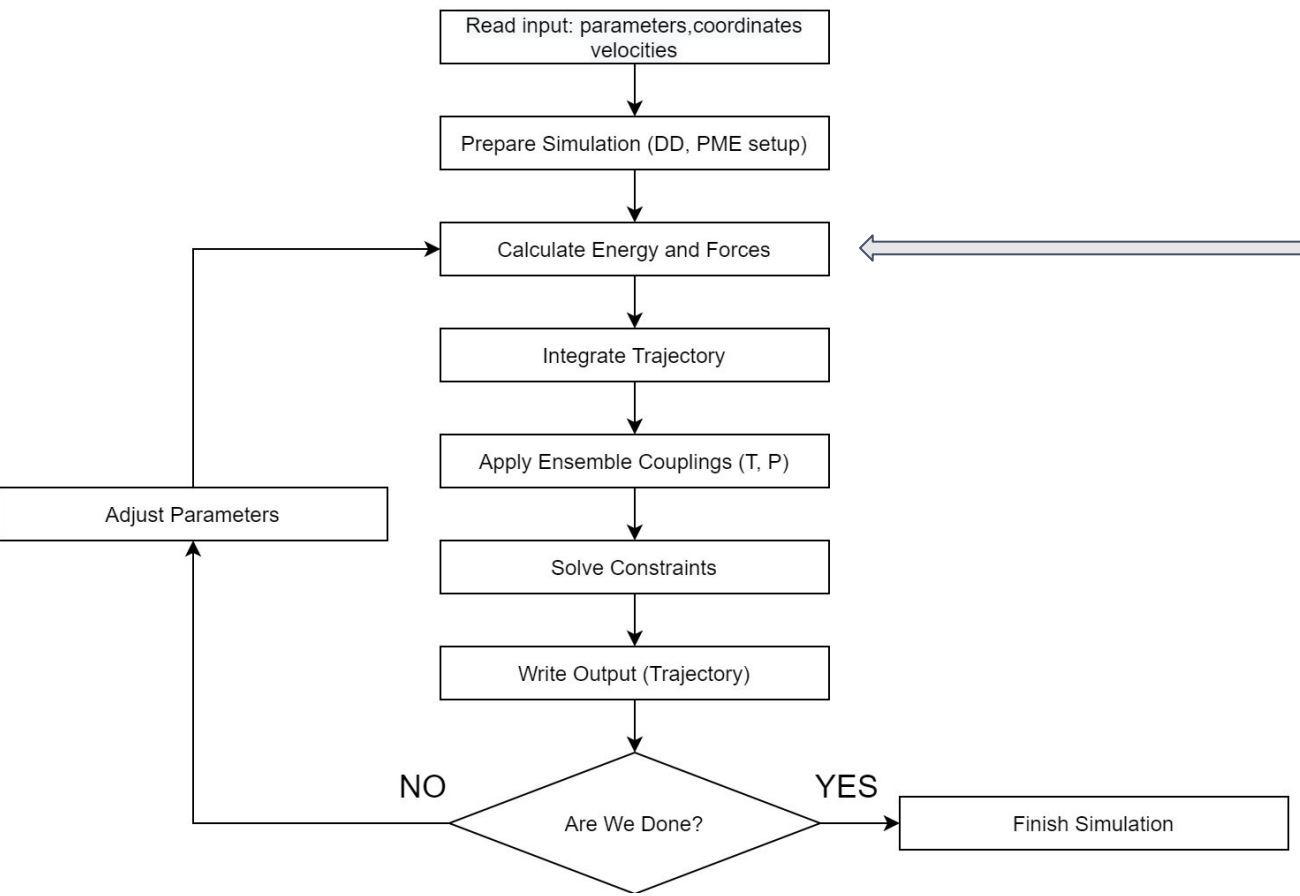
The University of Edinburgh

Horizon 2020 European Union Funding for Research & Innovation

Efficient (GROMACS+)CP2K compute resource usage for QM/MM simulation of biomolecular systems

- (GROMACS+)CP2K parallel execution
- The BioExcel QM/MM Benchmark Suite
- CP2K biomolecular QM/MM benchmark performance on CPUs
- CP2K biomolecular QM/MM benchmark performance on GPUs
- Summary and lessons learned - making efficient use of HPC compute resources

GROMACS+CP2K Parallel Execution



CP2K computes QM and QM-MM coupling energies and forces **in parallel** and passes these back to GROMACS

- QM and QM-MM calculations far more computationally costly than calculation of MM forces/energies and time integration
- Parallel performance of GROMACS+CP2K depends almost entirely on parallel performance of CP2K, given atomic/electronic properties received from GROMACS
- Need to understand how efficiently CP2K uses many CPU and GPU cores to decide how many compute resources to use when running GROMACS+CP2K (or CP2K standalone)

The BioExcel QM/MM Benchmark Suite

https://github.com/bioexcel/qmmm_benchmark_suite

Name	Type	Total atoms	QM atoms	Functional	QM Cell Size	Basis set	Time step
MQAE-BLYP	solute - solvent	~16 000	34	BLYP	14 x 17 x 11	DZVP-MOLOPT-GTH	1 fs
MQAE-B3LYP				B3LYP			
MQAE-B3LYP-large					28 x 34 x 22		
CBD_PHY-PBE	phytochrome	~168 000	68	PBE	25 x 25 x 25		
CBD_PHY-PBE0				PBE0			
ClC-19-BLYP	ion channel	~150 000	19	BLYP	18 x 18 x 18		
ClC-253-BLYP			253		27 x 25 x 25		

Adapted from *Extreme Scalability of DFT-Based QM/MM MD Simulations Using MiMiC. Viacheslav Bolnykh, Jógvan Magnus Haugaard Olsen, Simone Meloni, Martin P. Bircher, Emiliano Ippoliti, Paolo Carloni, and Ursula Rothlisberger. Journal of Chemical Theory and Computation 2019 15 (10), 5601-5613,* https://doi.org/10.1021/acs.jctc.9b00424

CPU Benchmarking Architectures & Protocol

CirrusCPU@EPCC (HPE SGI ICE XA):

- 2 x 18-core Intel Xeon Broadwell E5-2695@2.1 GHz
- 256GB RAM
- Infiniband
- 280 nodes / ~10k cores

Callum Bennetts/Maverick Photography

ARCHER2@EPCC (HPE CRAY EX):

- 2 x 64-core AMD EPYC Zen2 Rome 7742@2.25GHz
- 256GB RAM
- HPE Cray Slingshot
- 5860 nodes / ~750k cores

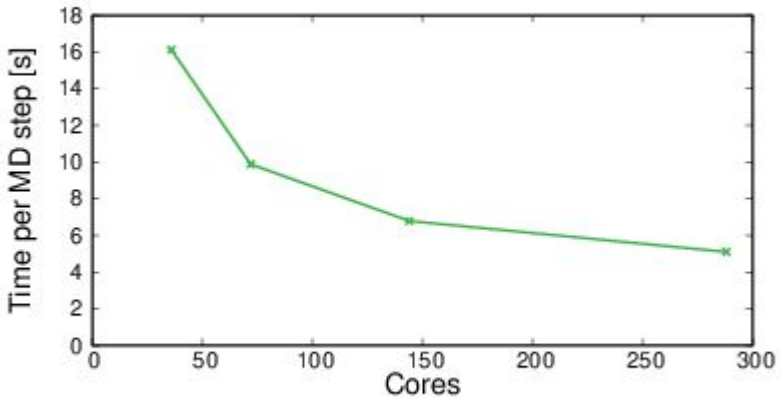
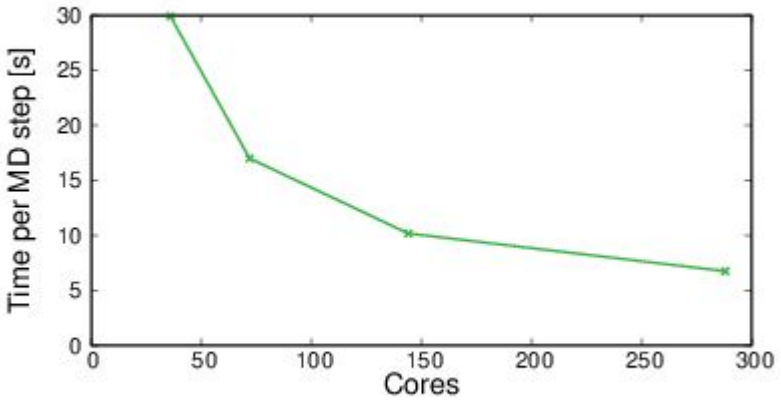
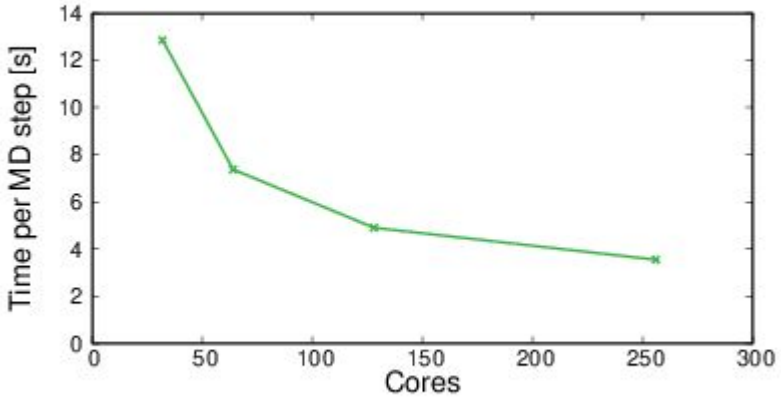
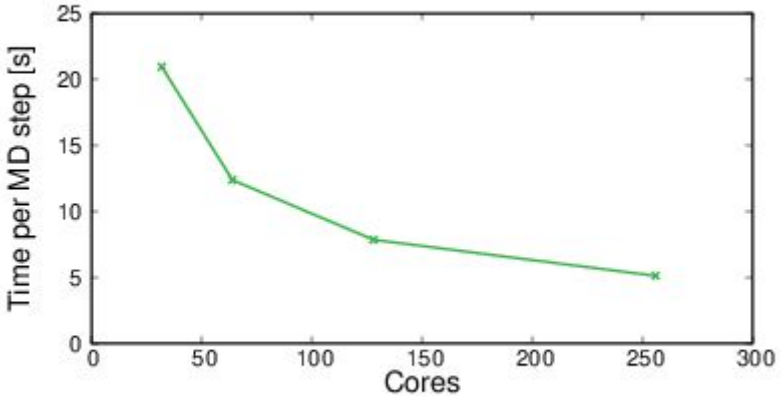
Protocol:

- Run 6 MD steps and subtract initialisation cost of 1 MD step
- Average over multiple runs
- Analyse using https://github.com/bioexcel/qmmm_benchmark_script

BioExcel QM/MM Benchmark Suite

Name	Type	Total atoms	QM atoms	Functional	QM Cell Size	Basis set	Time step
MQAE-BLYP	solute - solvent	~16 000	34	BLYP	14 x 17 x 11	DZVP-MOLOPT-GTH	1 fs
MQAE-B3LYP				B3LYP			
MQAE-B3LYP-large					28 x 34 x 22		
CBD_PHY-PBE	phytochrome	~168 000	68	PBE	25 x 25 x 25		
CBD_PHY-PBE0				PBE0			
ClC-19-BLYP	ion channel	~150 000	19	BLYP	18 x 18 x 18		
ClC-253-BLYP			253		27 x 25 x 25		

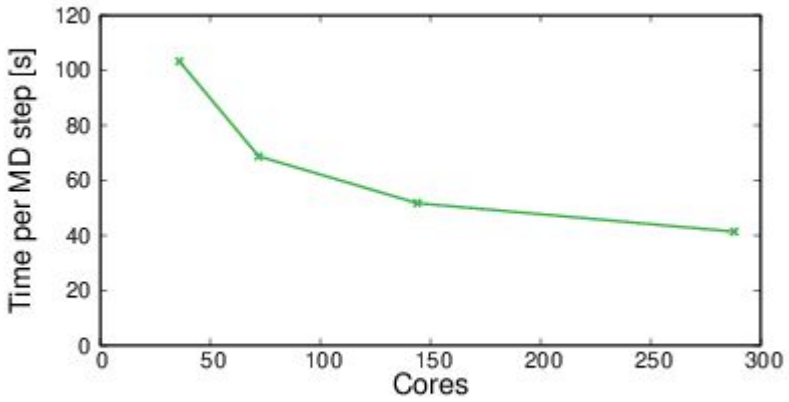
CPU benchmarking - MQAE

	MQAE-BLYP	MQAE-B3LYP
CirrusCPU		
	3 threads per rank Performance@288 cores: 17 ps/day Parallel efficiency@288 cores: 40%	6 threads per rank Performance@288 cores: 13 ps/day Parallel efficiency@288 cores: 55%
ARCHER2		
	4 threads per rank Performance@256 cores: 24 ps/day Parallel efficiency@256 cores: 45%	4 threads per rank Performance@256 cores: 17 ps/day Parallel efficiency@256 cores: 51%

CPU benchmarking - MQAE

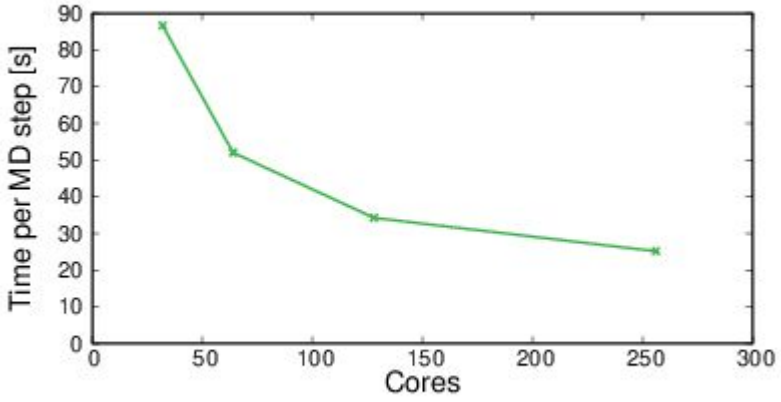
MQAE-B3LYP-large

CirrusCPU



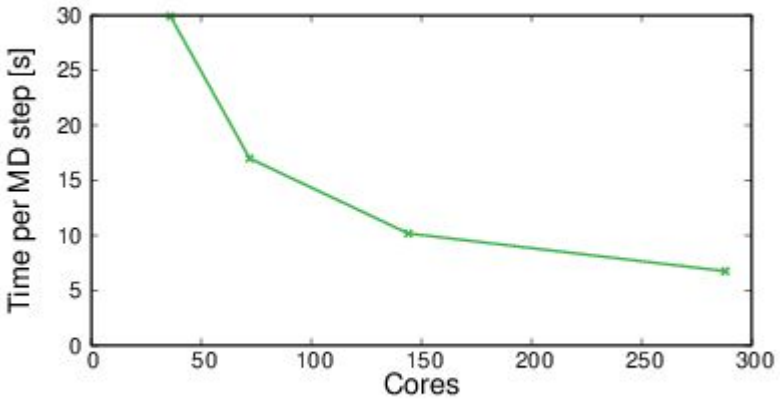
1 threads per rank
Performance@288 cores: 2.1 ps/day
Parallel efficiency@288 cores: 31%

ARCHER2

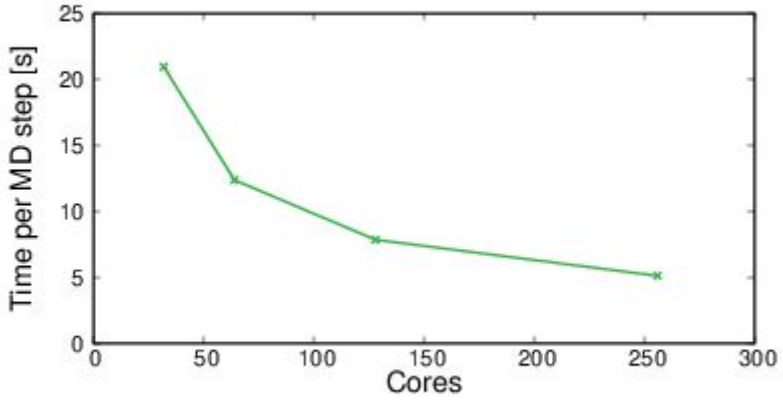


4 threads per rank
Performance@256 cores: 3.4 ps/day
Parallel efficiency@256 cores: 43%

MQAE-B3LYP



6 threads per rank
Performance@288 cores: 13 ps/day
Parallel efficiency@288 cores: 55%



4 threads per rank
Performance@256 cores: 17 ps/day
Parallel efficiency@256 cores: 51%

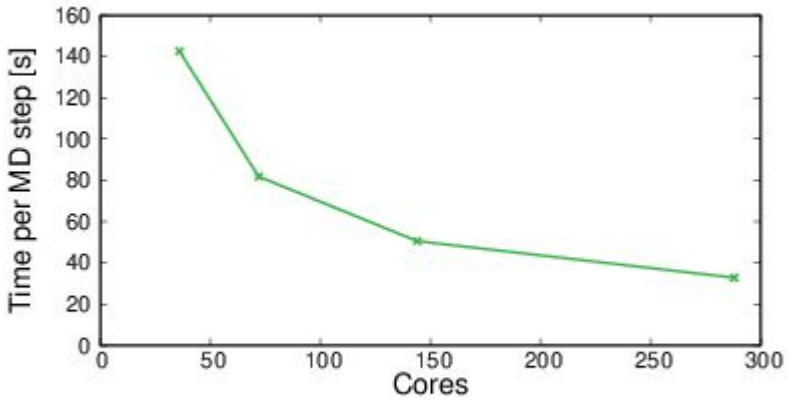
BioExcel QM/MM Benchmark Suite

Name	Type	Total atoms	QM atoms	Functional	QM Cell Size	Basis set	Time step
MQAE-BLYP	solute - solvent	~16 000	34	BLYP	14 x 17 x 11	DZVP-MOLOPT-GTH	1 fs
MQAE-B3LYP				B3LYP			
MQAE-B3LYP-large					28 x 34 x 22		
CBD_PHY-PBE	phytochrome	~168 000	68	PBE	25 x 25 x 25		
CBD_PHY-PBE0				PBE0			
ClC-19-BLYP	ion channel	~150 000	19	BLYP	18 x 18 x 18		
ClC-253-BLYP			253		27 x 25 x 25		

CPU benchmarking - CBD_PHY

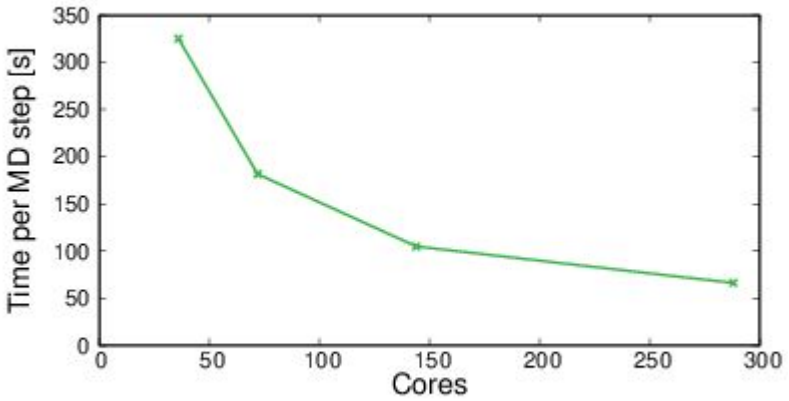
CBD_PHY-PBE

CirrusCPU



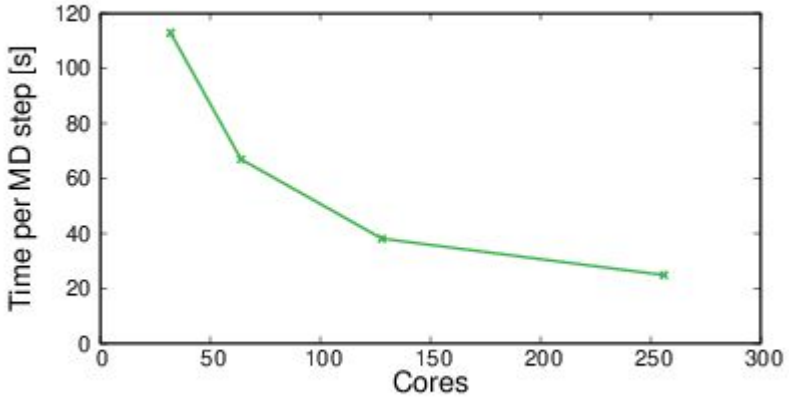
3 threads per rank
Performance@288 cores: 2.5 ps/day
Parallel efficiency@288 cores: 54%

CBD_PHY-PBE0

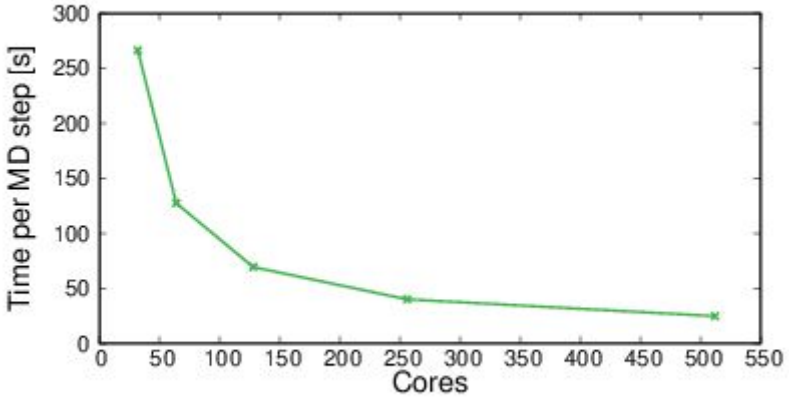


1 threads per rank
Performance@288 cores: 1.3 ps/day
Parallel efficiency@288 cores: 61%

ARCHER2



4 threads per rank
Performance@256 cores: 3.5 ps/day
Parallel efficiency@256 cores: 56%



4 threads per rank
Performance@256 cores: 2.1 ps/day
Parallel efficiency@256 cores: 83%

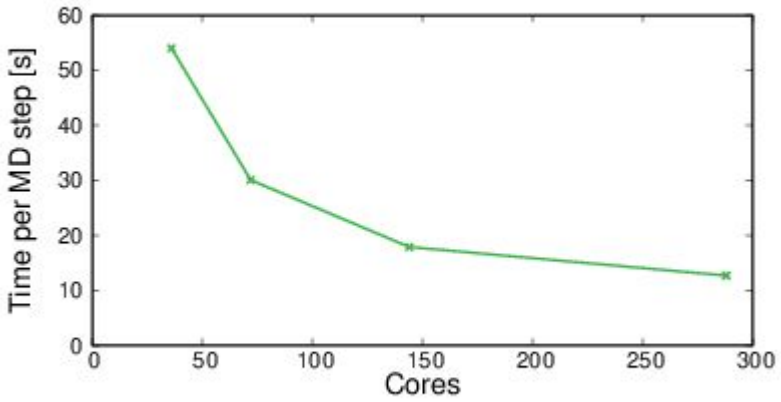
BioExcel QM/MM Benchmark Suite

Name	Type	Total atoms	QM atoms	Functional	QM Cell Size	Basis set	Time step
MQAE-BLYP	solute - solvent	~16 000	34	BLYP	14 x 17 x 11	DZVP-MOLOPT-GTH	1 fs
MQAE-B3LYP				B3LYP			
MQAE-B3LYP-large					28 x 34 x 22		
CBD_PHY-PBE	phytochrome	~168 000	68	PBE	25 x 25 x 25		
CBD_PHY-PBE0				PBE0			
ClC-19-BLYP	ion channel	~150 000	19	BLYP	18 x 18 x 18		
ClC-253-BLYP			253		27 x 25 x 25		

CPU benchmarking - ClC

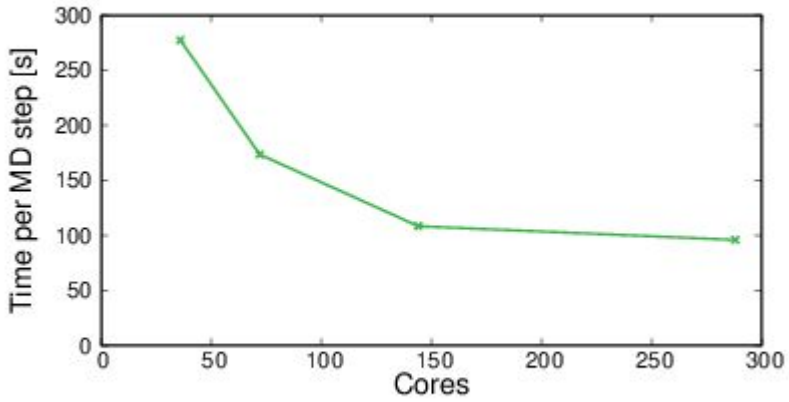
ClC-19-BLYP

CirrusCPU



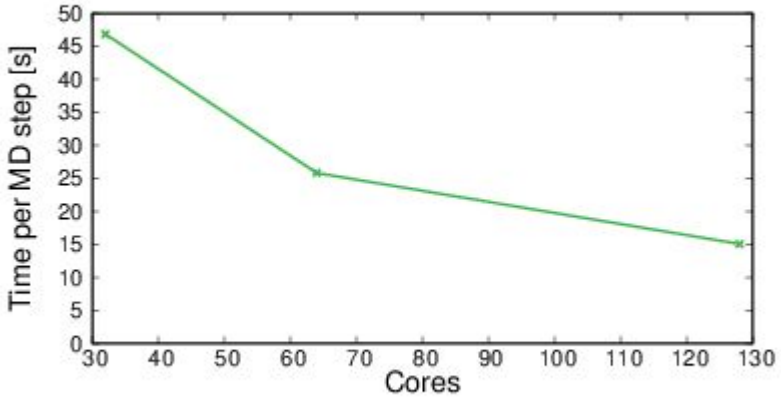
3 threads per rank
Performance@288 cores: 6.8 ps/day
Parallel efficiency@288 cores: 53%

ClC-253-BLYP

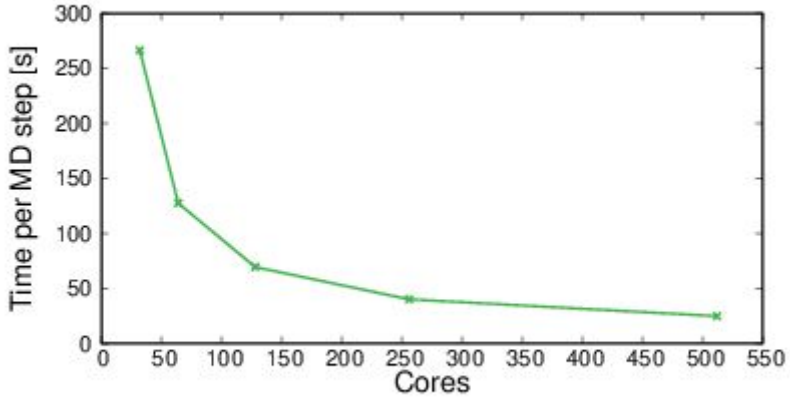


6 threads per rank
Performance@288 cores: 1 ps/day
Parallel efficiency@288 cores: 36%

ARCHER2



4 threads per rank
Performance@128 cores: 3.8 ps/day
Parallel efficiency@128 cores: 58%



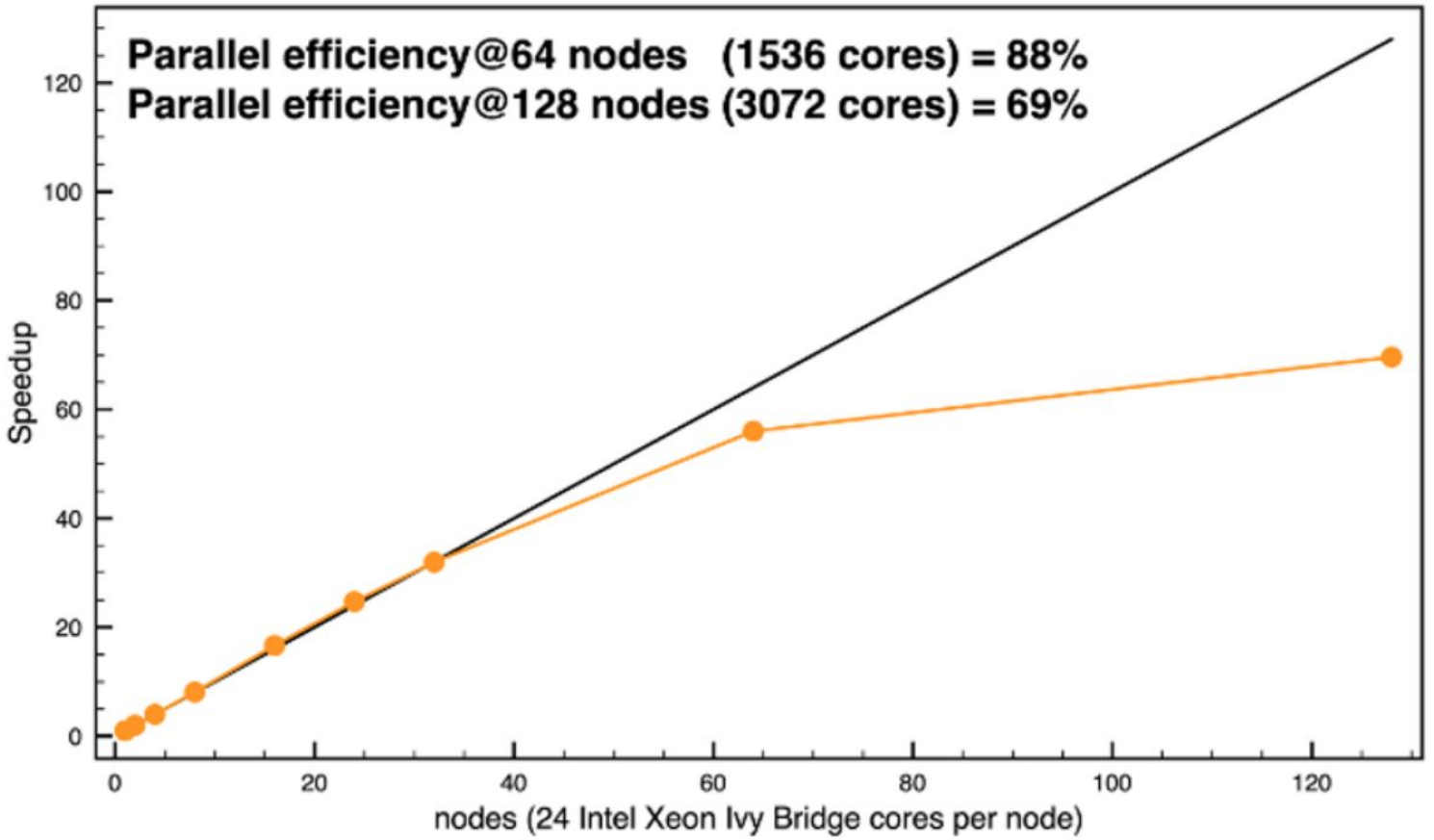
6 threads per rank
Performance@256 cores: 0.7 ps/day
Parallel efficiency@256 cores: 21%

A helping hand for hybrid functional calculations

- MOLOPT basis set very costly combined with hybrid functionals (HFX)
- Auxiliary Density Matrix Method (ADMM) can help:
  - e.g. MOLOPT+ADMM 15x faster for ClC-19-B3LYP than standard MOLOPT
- May consider other basis sets: EMSL, HFX_BASIS, but may be less suitable for biomolecular systems
  - → check the literature

Beyond Hartree-Fock Exchange

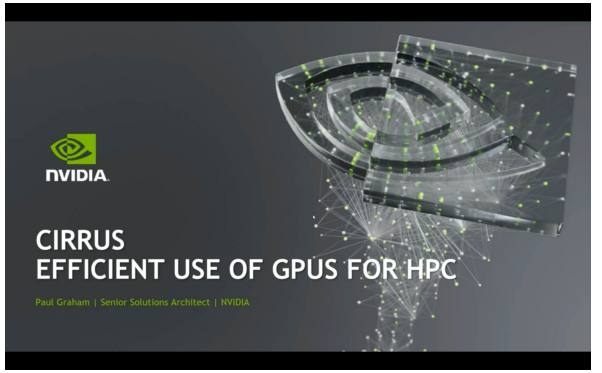
CBD_PHY-MP2



GPU Benchmarking Architectures

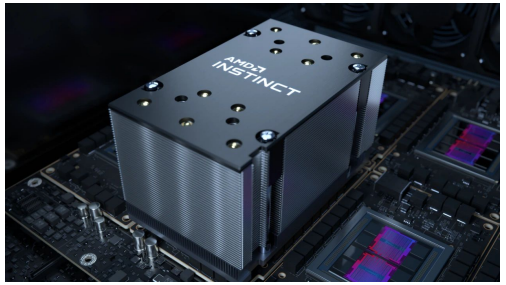
CirrusGPU@EPCC (HPE SGI ICE XA):

- 4 x NVIDIA Tesla V100 (Volta) SXM2-16GB
- 2 x 20-core Intel Xeon (Cascade Lake) Gold 6248@2.4 GHz
- 384GB RAM
- Infiniband



AMD Accelerator Cloud

- 1 x 64-core AMD EPYC + 8 x AMD Instinct MI100 GPU
- 512GB RAM
- Infiniband



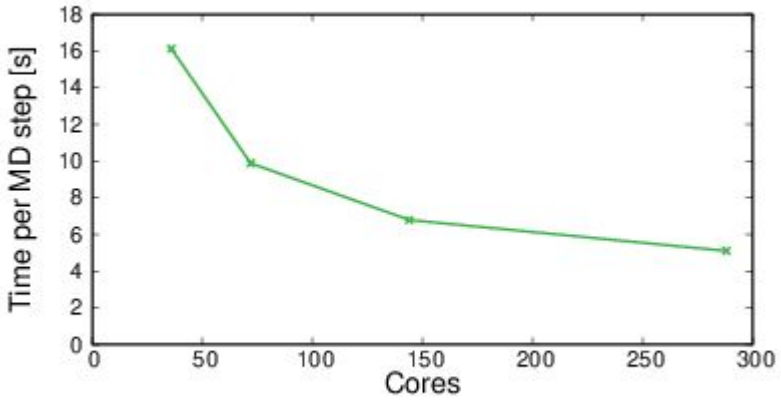
BioExcel QM/MM Benchmark Suite

Name	Type	Total atoms	QM atoms	Functional	QM Cell Size	Basis set	Time step
MQAE-BLYP	solute - solvent	~16 000	34	BLYP	14 x 17 x 11	DZVP-MOLOPT-GTH	1 fs
MQAE-B3LYP				B3LYP			
MQAE-B3LYP-large					28 x 34 x 22		
CBD_PHY-PBE	phytochrome	~168 000	68	PBE	25 x 25 x 25		
CBD_PHY-PBE0				PBE0			
ClC-19-BLYP	ion channel	~150 000	19	BLYP	18 x 18 x 18		
ClC-253-BLYP			253		27 x 25 x 25		

NVIDIA GPU benchmarking - MQAE

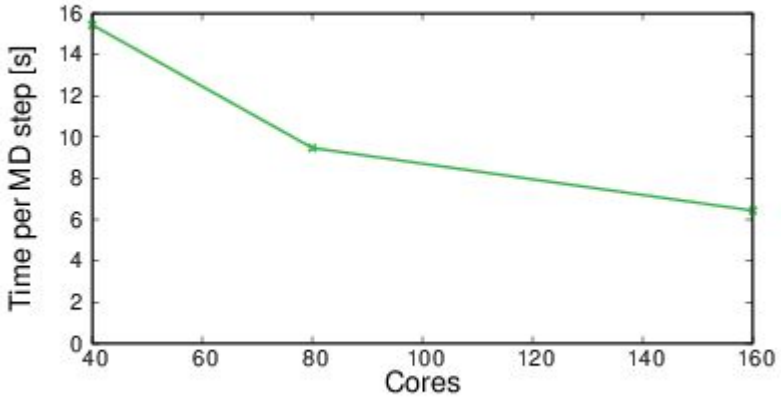
MQAE-BLYP

CirrusCPU



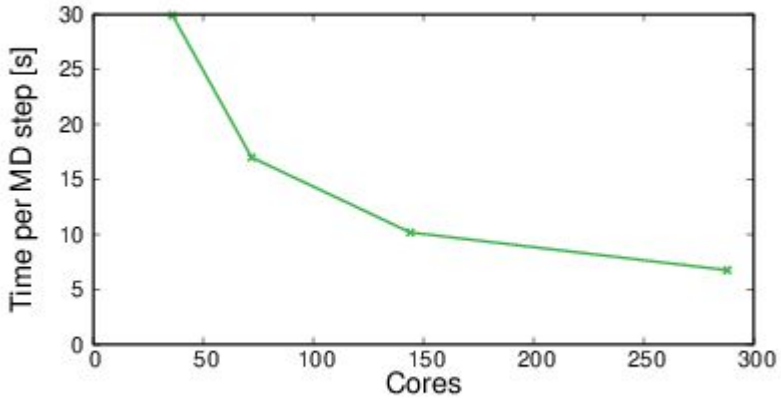
3 threads per rank
Performance@144 cores: 12.7 ps/day
Parallel efficiency@144 cores: 59%

CirrusGPU

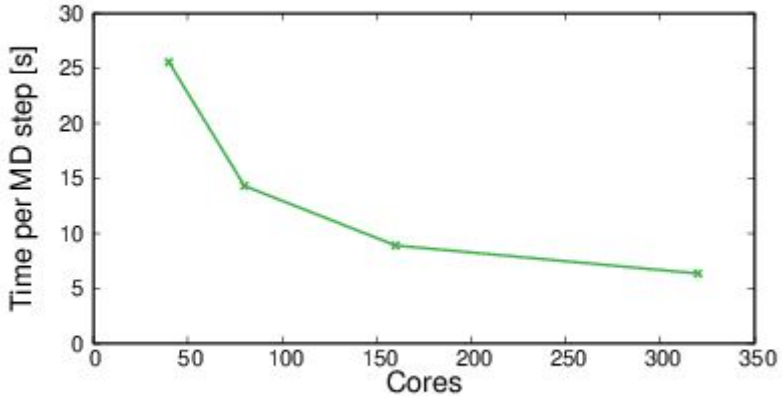


10 threads per rank
Performance@160 cores: 13.4 ps/day
Parallel efficiency@160 cores: 60%

MQAE-B3LYP



6 threads per rank
Performance@144 cores: 8.5 ps/day
Parallel efficiency@144 cores: 73%

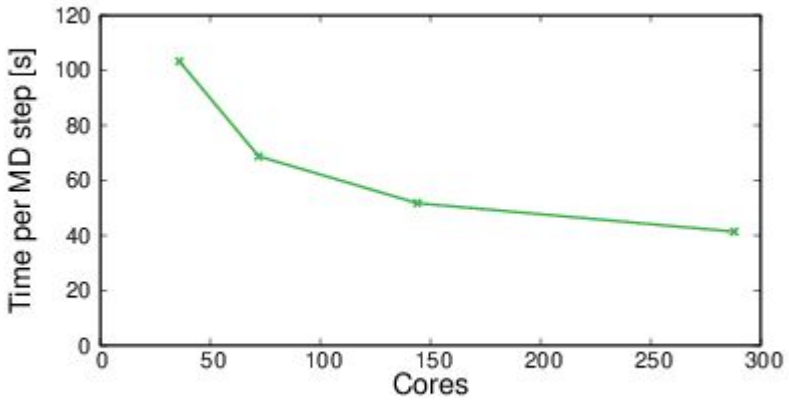


10 threads per rank
Performance@160 cores: 9.7 ps/day
Parallel efficiency@160 cores: 71%

NVIDIA GPU benchmarking - MQAE

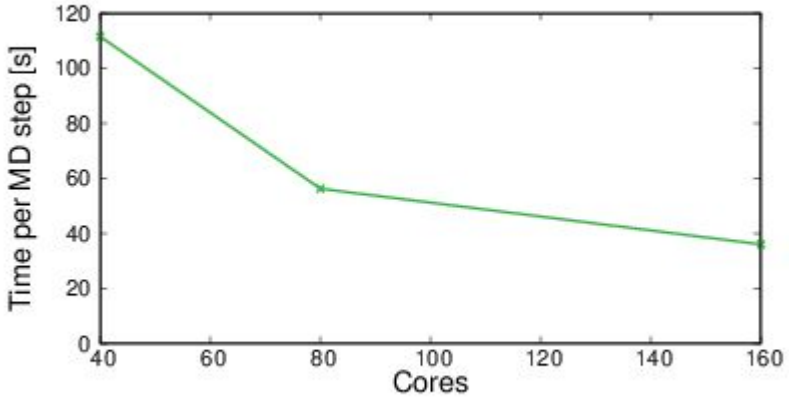
MQAE-B3LYP-large

CirrusCPU



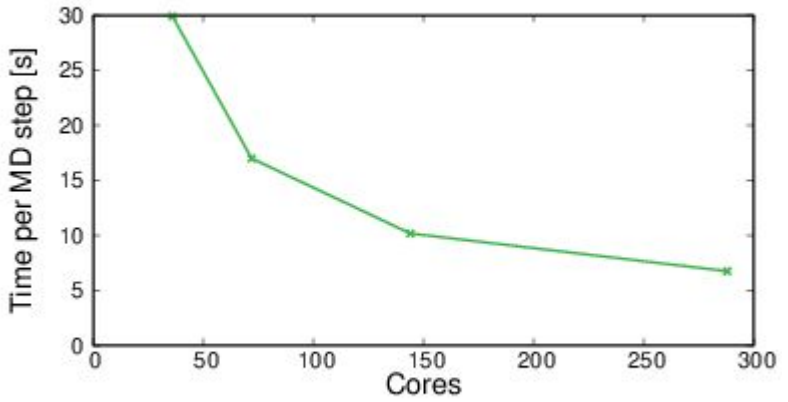
1 threads per rank
Performance@144 cores: 17 ps/day
Parallel efficiency@144 cores: 50%

CirrusGPU

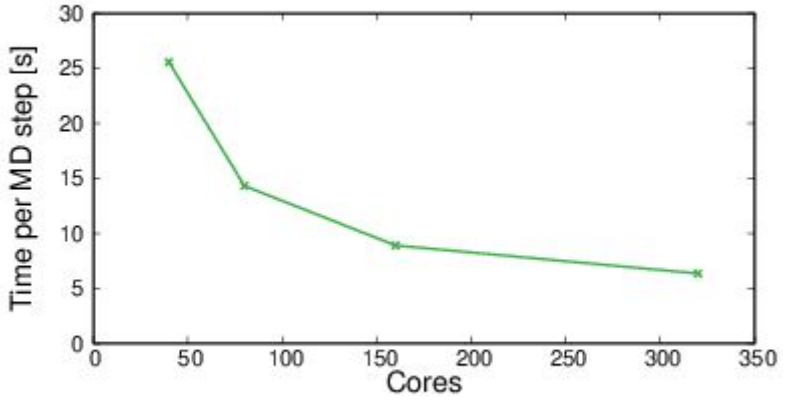


10 threads per rank
Performance@160 cores: 24 ps/day
Parallel efficiency@160 cores: 77%

MQAE-B3LYP



6 threads per rank
Performance@144 cores: 8.5 ps/day



10 threads per rank
Performance@160 cores: 9.7 ps/day
Parallel efficiency@160 cores: 71%

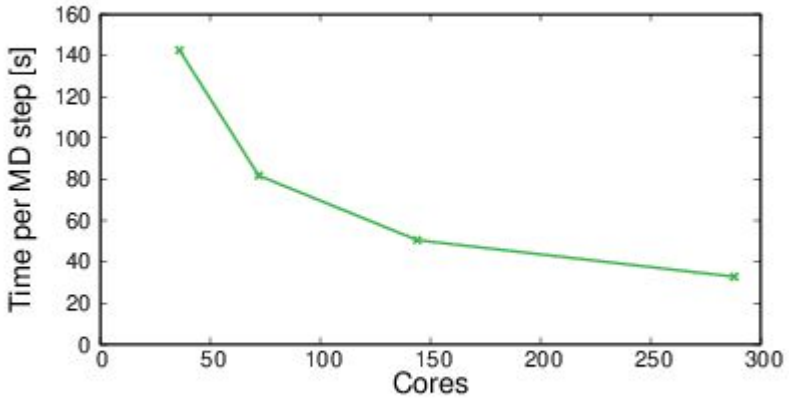
BioExcel QM/MM Benchmark Suite

Name	Type	Total atoms	QM atoms	Functional	QM Cell Size	Basis set	Time step
MQAE-BLYP	solute - solvent	~16 000	34	BLYP	14 x 17 x 11	DZVP-MOLOPT-GTH	1 fs
MQAE-B3LYP				B3LYP			
MQAE-B3LYP-large					28 x 34 x 22		
CBD_PHY-PBE	phytochrome	~168 000	68	PBE	25 x 25 x 25		
CBD_PHY-PBE0				PBE0			
ClC-19-BLYP	ion channel	~150 000	19	BLYP	18 x 18 x 18		
ClC-253-BLYP			253		27 x 25 x 25		

NVIDIA GPU benchmarking - CBD_PHY

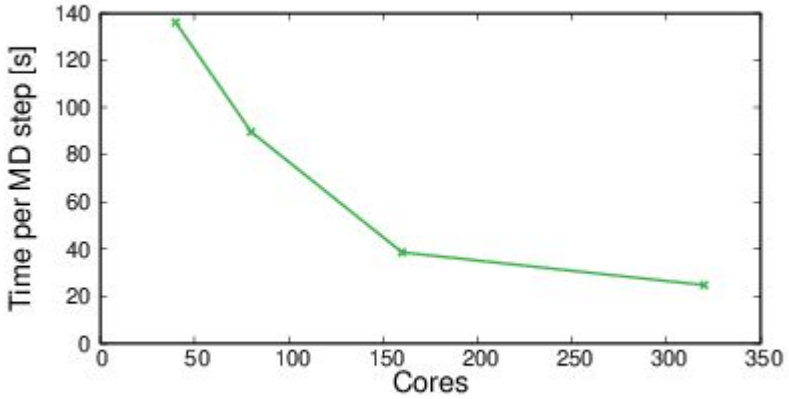
CBD_PHY-PBE

CirrusCPU



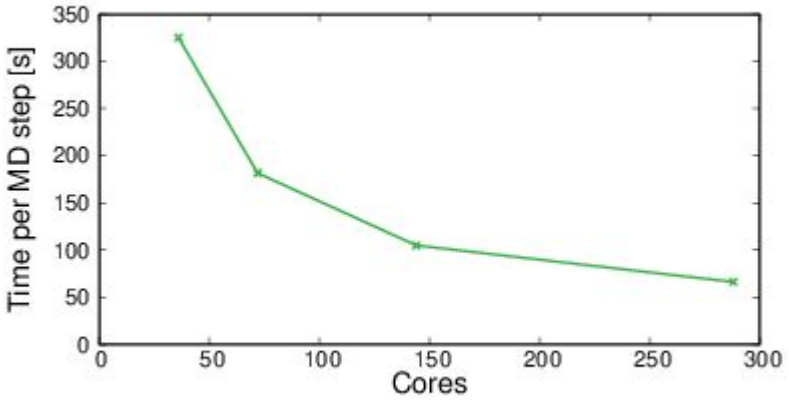
3 threads per rank
Performance@288 cores: 2.5 ps/day
Parallel efficiency@288 cores: 54%

CirrusGPU

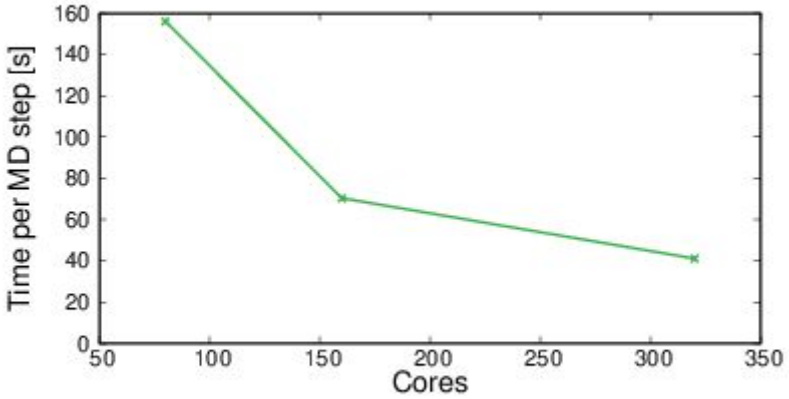


10 threads per rank
Performance@320 cores: 3.5 ps/day
Parallel efficiency@320 cores: 69%

CBD_PHY-PBE0



1 threads per rank
Performance@288 cores: 1.3 ps/day
Parallel efficiency@288 cores: 61%



10 threads per rank
Performance@320 cores: 2.1 ps/day

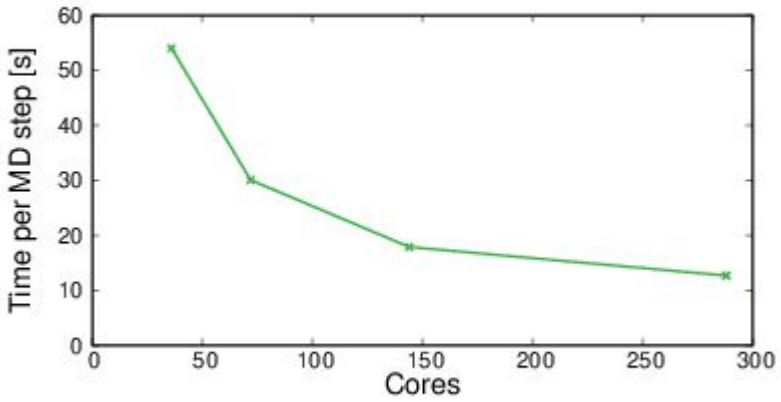
BioExcel QM/MM Benchmark Suite

Name	Type	Total atoms	QM atoms	Functional	QM Cell Size	Basis set	Time step
MQAE-BLYP	solute - solvent	~16 000	34	BLYP	14 x 17 x 11	DZVP-MOLOPT-GTH	1 fs
MQAE-B3LYP				B3LYP			
MQAE-B3LYP-large					28 x 34 x 22		
CBD_PHY-PBE	phytochrome	~168 000	68	PBE	25 x 25 x 25		
CBD_PHY-PBE0				PBE0			
ClC-19-BLYP	ion channel	~150 000	19	BLYP	18 x 18 x 18		
ClC-253-BLYP			253		27 x 25 x 25		

NVIDIA GPU benchmarking - ClC

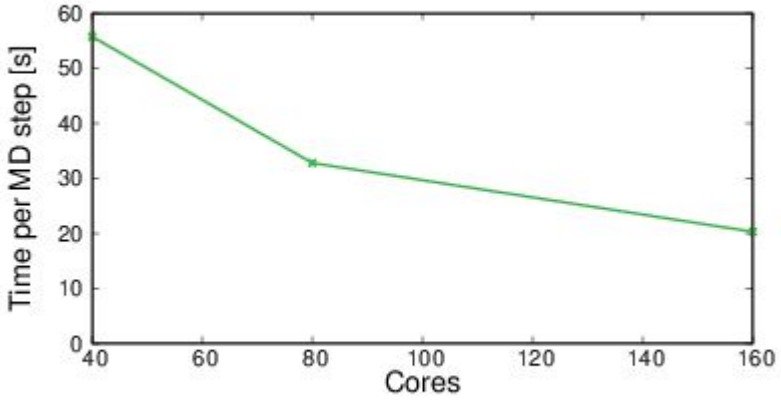
ClC-19-BLYP

CirrusCPU



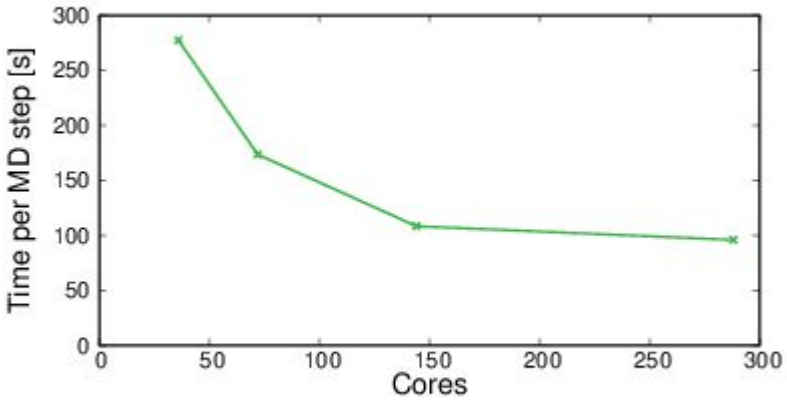
3 threads per rank
Performance@144 cores: ps/day
Performance@288 cores: 6.8 ps/day
Parallel efficiency@288 cores: 53%

CirrusGPU

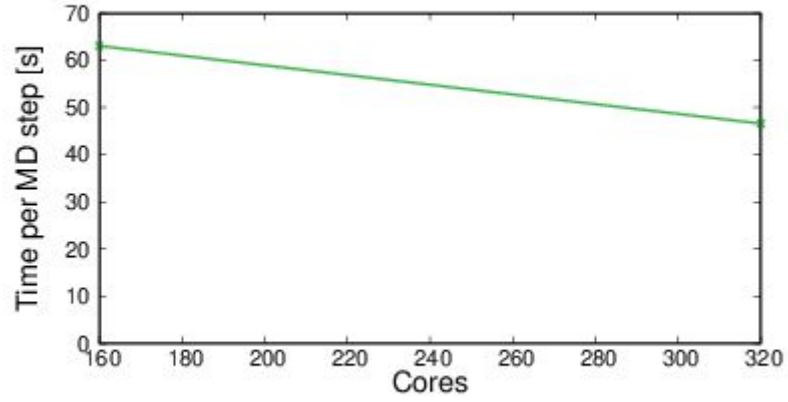


10 threads per rank
Performance@160 cores: 4.3 ps/day
Parallel efficiency@160 cores: 69%

ClC-253-BLYP



6 threads per rank
Performance@144 cores: 0.8 ps/day
Performance@288 cores: 1 ps/day



10 threads per rank
Performance@160 cores: 1.4 ps/day
Performance@320 cores: 1.9 ps/day

NVIDIA GPU performance summary

- Typical QM treatments for biomolecular QM/MM simulation, especially small number of QM atoms, do not heavily rely on parts of CP2K or external libraries that benefit strongly from GPU offloading

- Very active ongoing GPU offloading development within CP2K is due to improve this

Status of GPU Support in CP2K

Library	Status	Accelerates	Backends
DBCSR	Ready	LS-SCF	CUDA, HIP, OpenCL
grid	Ready	GPW	CUDA, HIP
pw	Ready	SCCS	CUDA, HIP
COSMA	Ready	RPA	CUDA, HIP
SIRIUS	Ready	PW DFT	CUDA, HIP
ELPA	Ready (kinda)	Diagonalization	CUDA
Alternative eigensolver	In progress	Diagonalization	CUDA
SpFFT	In progress	GPW, SCCS	CUDA, HIP
DBM	In progress	RI methods	CUDA, HIP
SPLA	In progress	MP2	CUDA, HIP
Two-electron integrals	In progress	HFX	
GEEP	Planned	QM/MM	
libxc	Planned	GPW	
One-electron integrals	Planned	GPW	

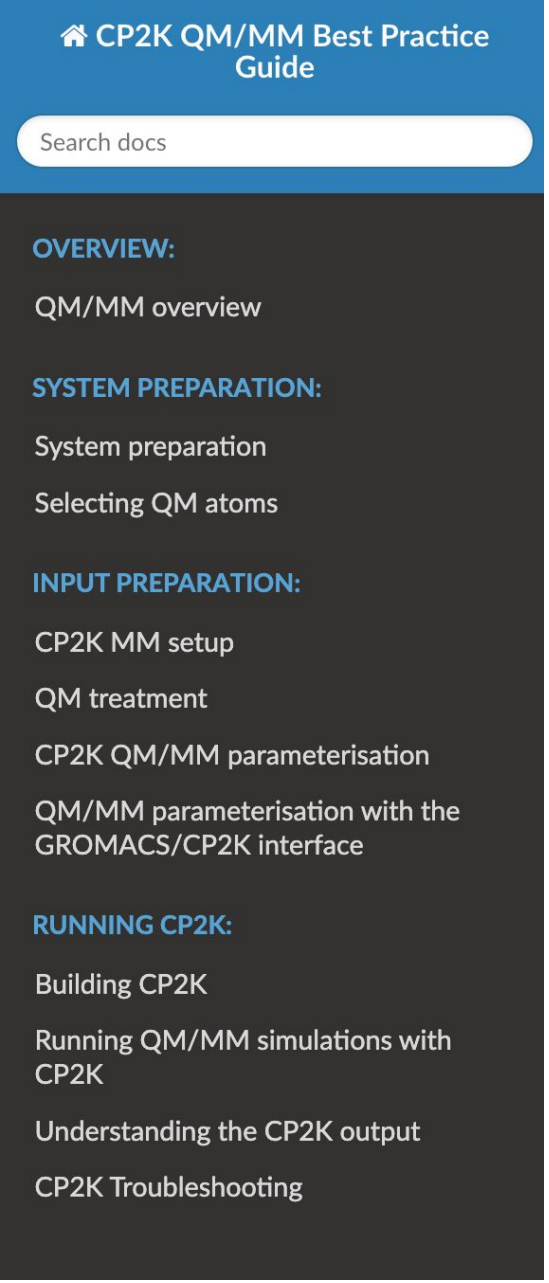
CP2K AMD GPU support

CP2K v9.1 has (experimental) HIP offload support:

- DBCSR HIP offloading
  - also uses hipBLAS
  - Issue involving OpenMP multithreading
- HIP backend can offload grid operations to AMD GPUs
- ELPA library has work-in-progress HIP support
  - Not recommended for production use (yet)
- COSMA library supports rocBLAS/hipBLAS
  - Not recommended for production use (yet)

Performance:

- Single node benchmarking of MQAE-BLYP
- 8 ranks (1 per MI100 GPU)
- 8 threads per rank
- 11s per MD step
  - single node CirrusGPU (40 Intel Cascade Lake cores + 4 x NVIDIA V100: 15.4s per MD step)
  - half node ARCHER2 (64 AMD EPYC cores: 7.3s per MD step)



What next?

- Follow our (GROMACS+)CP2K Best Practice Guide: https://docs.bioexcel.eu/qmmm_bpg
- Find support using GROMACS+CP2K for biomolecular QM/MM: https://ask.bioexcel.eu
- Visit our workshop on Best Practices in QM/MM simulation of biomolecular systems: https://bioexcel.eu/events/virtual-workshop-best-practices-in-qm-mm-simulation-of-biomolecular-systems/
- Follow our webinar about the GROMACS-CP2K interface: https://bioexcel.eu/webinar-multiscale-qm-mm-simulations-exploring-chemical-reactions-using-novel-gromacs-cp2k-interface-2020-12-08/

Thanks







Audience Q&A session

- Please use the Q&A function at the bottom of **Zoom** application

- Any other questions or points to discuss after the webinar?
  Join the discussions at http://ask.bioexcel.eu

BioExcel Partners

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

NORMAN CONSULTING

EMBL-EBI

KTH VETENSKAP OCH KONST

THE UNIVERSITY OF EDINBURGH

JÜLICH FORSCHUNGSZENTRUM

IRB BARCELONA
INSTITUTO DE INVESTIGACIÓN BIOMÉDICA

Utrecht University

NBD NOSTRUM BIODISCOVERY

acrosslimits

Ian Harrow Consulting

MAX-PLANCK-GESELLSCHAFT

MANCHESTER 1824

UNIVERSITY OF JYVÄSKYLA

Horizon 2020
European Union Funding
for Research & Innovation

BioExcel is funded by the European Union Horizon 2020 program under grant agreements 675728 and 823830.