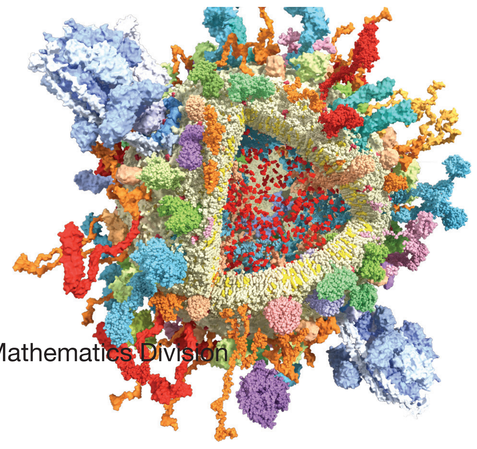
Unified long-range electrostatics and dynamic protonation for realistic biomolecular simulations on the Exascale



Applicants: Helmut Grubmüller,[1] Holger Dachsel,[2] Berk Hess[3]
Other Participating individuals: Ivo Kabadschow,[2] Carsten Kutzner[1]

[1]Theoretical and computational Biophysics, MPI for biophysical Chemistry, Göttingen, [2]Jülich Supercomputing Centre, Institute for Advanced Simulation, Mathematics Division
[3]Dept. of Theoretical Physics, School of Engineering Sciences, RIT, Stockholm

Abstract

- We target a flexible and portable exascale algorithm for potentials and forces, a prerequisite for exascale applications in particle-based simulations with long-range interactions in general (Fig. 1).
- To develop a general tool box for a wide range of exascale applications, we will perform an in-depth analysis of a well chosen real-world example:
- The molecular dynamics (MD) simulation software GROMACS[1] is a prototypic example of a general class of complex multiparticle algorithms.
- MD is challenging not only because of the **long-range effects of the electrostatic interactions** but also because the **location of partial charges is time-dependent**, as the protonation states of the molecules vary rapidly (Fig. 5).

In this proposal we address both tighly interlinked challenges by the development of a **flexible, unified, ultra-scalable toolkit for long-range interactions** that will as a direct outcome provide realistic, dynamic protonation states in exascale MD applications. Moreover, it will provide a toolbox and a framework for future scientists who need to build their own exa-scalable applications.

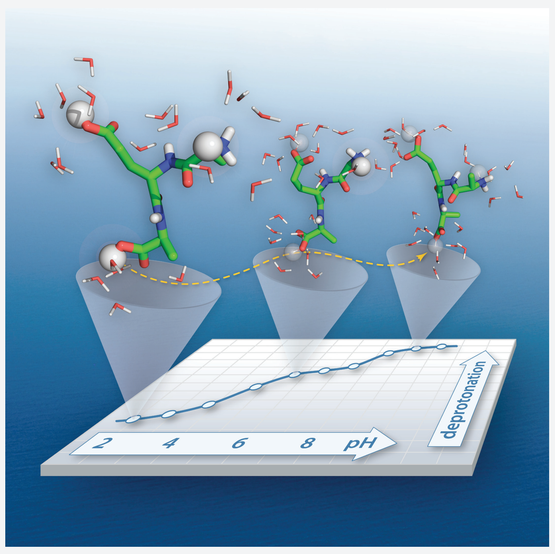
State of the art

Strongly interacting multiparticle systems lead to two lines of challenges:

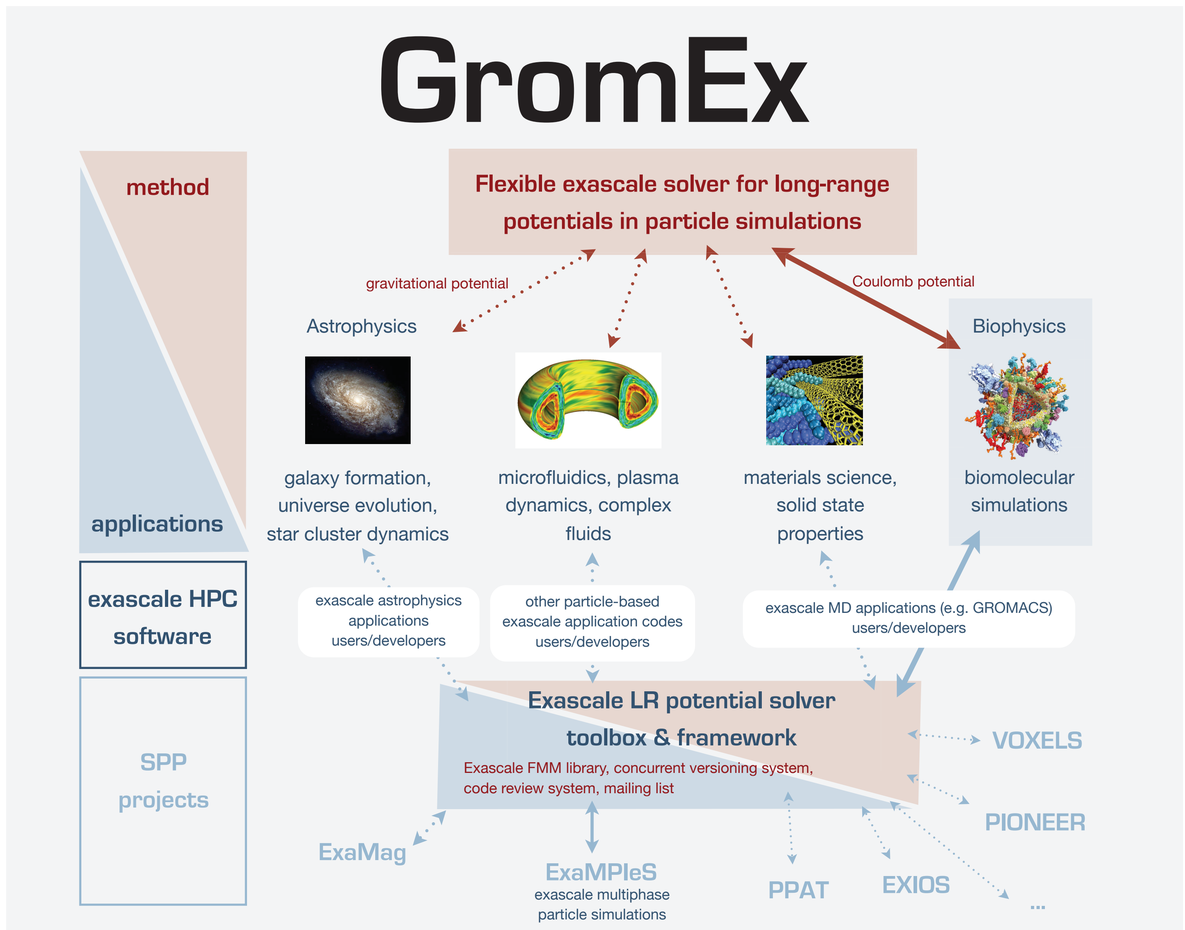
- First, the inter-particle interactions are typically long-range, which calls for numerical techniques that cut down the **O(N²) complexity,** ideally to O(N). Different schemes have been developed and applied, such as Particle Mesh Ewald, all facing communication bottlenecks[2] and thus scalability problems on large parallel machines.
- The second ubiquitous challenge is that of **time-dependent interaction functions**, i.e., changing masses of the interacting galaxies or stars, changing ionization states in plasmas, or **changing titration states** in fluids or macromolecules,[3] Fig. 5. This second challenge poses severe communication bottlenecks, thus calling for sophisticated computation/communication balancing algorithms.

Dynamic protonation

- The electrostatic potential and the location of charges on the protein are tightly coupled
- The pH-dependent protonation continuously changes in response to the electrostatic environment.
- Particularly for larger functional, conformational, or electrostatic changes, the lack of appropriate algorithms for this process poses a severe limitation for realistic biomolecular simulations.
- Recently, the λ-dynamics scheme has been designed by one of the applicants that allows for a dynamic protonation protocol, thereby allowing for realistic constant-pH simulations.[3]
- The constant-pH concept requires the consideration of multiple locally differing charge distributions.
- PME, however, requires the separate evaluation of the Fourier coefficients for each charge alternative. Each of these evaluations involves expensive all-to-all communication which severly limits the parallel scaling.



↗ **Fig. 5: For realistic biomolecular simulations, dynamic protonation is essential.[3] The electrostatic solver must be able to efficiently handle various protonation states.**



↗ **Fig. 1: The GromEx project proposes a flexible solver that is needed for all kinds of particle-based parallel simulations (top) plus its realization in an exascale library framework (bottom). The building blocks of the exascale library can be used (and enhanced) by developers of application codes ranging from astrophysics to atomistic simulations.**

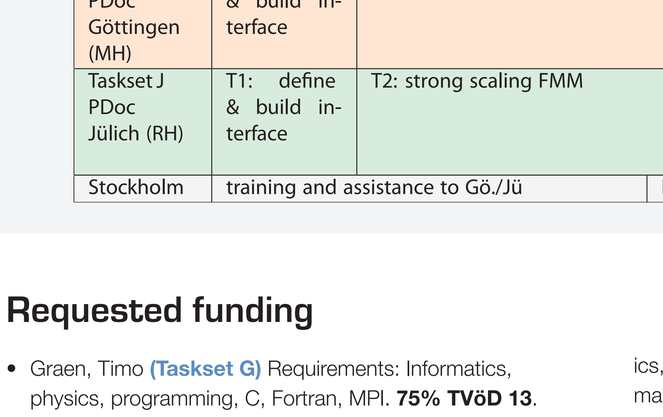
↘ **Table 1: Estimate of distribution of work-packages/tasks T1–T5 by half-year and group.**

Year	2013		2014		2015	
Taskset G PhD Göttingen (TG)	T1: benchmarks	T4: enhance FMM for alternative particle properties				T5: showcase application(s)
Taskset H PDoc Göttingen (MH)	T1: define & build interface	T3: alternative particle properties			T3: correctness tests T5: set up collaboration framework	
Taskset J PDoc Jülich (RH)	T1: define & build interface	T2: strong scaling FMM			T4: adapt FMM far field to local changes	T5: Documentation
Stockholm	training and assistance to Gö./Jü			Incorporation in GROMACS, help on T5		

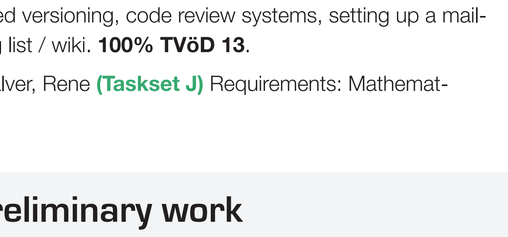
Requested funding

- Graen, Timo **(Taskset G)** Requirements: Informatics, physics, programming, C, Fortran, MPI. **75% TVöD 13**.
- Hoefling, Martin **(Taskset H)** Requirements: Physics, biomolecular simulation, IT, programming C, MPI, OpenMP, web-based application administration, distributed versioning, code review systems, setting up a mailing list / wiki. **100% TVöD 13**.
- Halver, Rene **(Taskset J)** Requirements: Mathematics, programming C and Fortran, MPI, OpenMP, performance monitoring and optimization of extreme-scale parallel applications. **100% TVöD 13**.
- Stockholm is not asking for own funding, but will provide training and support for TG, MH, and RH throughout the project with workshops and internships.
- Travel expenses: **5000 € / year**
- Publications (open access): **2000 € / year**

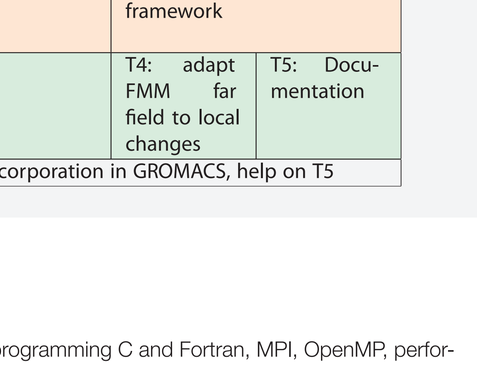
Preliminary work

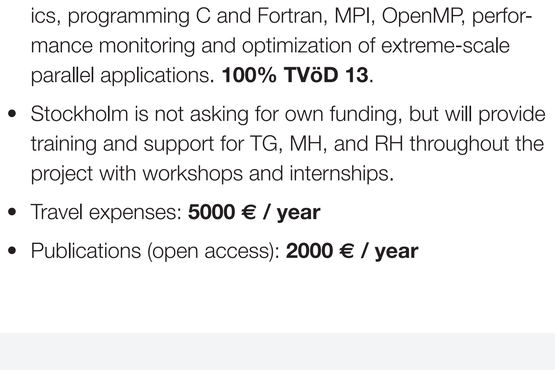


↗ **Fig. 2: FMM scaling on Juropa (Intel 3 GHz) with relative error in energy of 10^{-3} using 1 M particles.**



↗ **Fig. 3: GROMACS 4.5 scaling on 3 different HPC clusters at the Garching supercomputing centre.**





↗ **Fig. 4 Top: FMM scaling on Juropa with relative error in energy of 10^{-3} using 4096 to 16 M particles. Bottom: GROMACS time step lengths on a single core.**

Objectives & Work Programme

Develop an ultra-scalable, versatile solver for complex long-range interactions applicable to particle-based simulations in general.
The main goal concerning the derivation of forces from 1/r-type potentials will be (a) to optimize the scalability and (b) to broaden the applicability to the general situations and geometries needed by all types of particle-based applications. To reach these ambitious goals, we will

- characterize the scaling behavior of different long range methods under exascale conditions,
- in a close co-design loop, identify the method that proves most generalizable – likely the fast multipole method (FMM) (Fig. 2) and structure-adapted FMM[4] (Fig. 6),
- generate a toolbox of how to advance the FMM to the exascale level for various testsystems,
- generate a toolbox of how to overcome the communication bottleneck for time-dependent interactions,
- evaluate the toolbox by sample implementations, e.g., biomolecular and astrophysical simulations, both for (a) long range FMM and (b) dynamic charge/mass changes.

The project consists of 5 tasks, of which tasks 2–4 can be started independently, while all tasks involve a very close collaboration between the groups (Table 1).

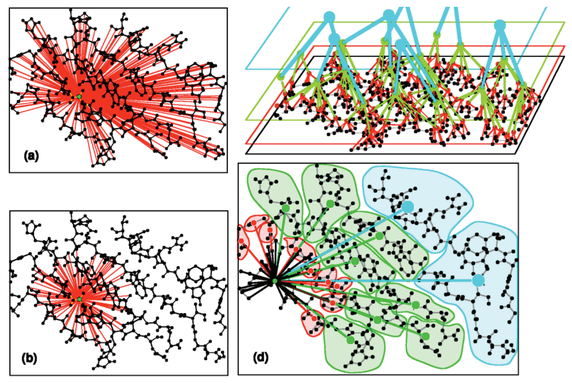
Task 1: Evaluation and benchmark phase (shared among **TG**, **MH** and **RH**). With the help of the Jülich *Scalable Fast Coulomb Solver*,[5–7] various solvers for the long range potential will be evaluated for their exascale potential.

Task 2: Strong scaling FMM (**RH**). Redesign FMM for exascale application using one-sided, point-to-point communication, overlap communication and computation, further allow for local updates of positions/charges without updates of the far field contributions. Remove synchronization points in the tree hierarchy, keep sorting local.

Task 3: Allow for alternative particle properties in the MD engine (**MH**). In the constant pH case, multiple charge distributions have to be evaluated, generating a force on the virtual λ particle, which in turn defines the current protonation state of the residue.

Task 4: Advance the FMM to allow efficient calculation of locally alternative charge distributions for which the potentials and forces are communicated back to the main simulation engine. (**TG**, with the help of **RH**)

Task 5: Integration, dissemination, and showcase application (all groups)



↗ **Fig. 6 The structure-adapted FMM.[4]**

Project-related publications

1 B Hess, C Kutzner, D vd Spoel, and E Lindahl. ***JCTC***, 2008

2 C Kutzner, D vd Spoel, M Fechner, E Lindahl, UW Schmitt, BL de Groot, and H Grubmüller. ***JCC***, 2007

3 S Donnini, F Tegeler, G Groenhof, and H Grubmüller. ***JCTC***, 2011

4 M. Eichinger, H. Grubmüller, H. Heller, P. Tavan ***JCC***, 1997

5 H Dachsel. ***JCP***, 2006

6 H Dachsel. ***JCP***, 2010

7 I Kabadshow and H Dachsel. ***IAS series***, Fast Methods for Long-Range Interactions in Complex Systems, 2011

8 S Pronk, P Larsson, B Hess, ..., VS Pande, P Kasson, and E Lindahl. ***Proceedings of SC'2011***

9 C Kutzner, H Grubmüller, BL de Groot, and U Zachariae. ***BJ***, 2011

10 H Grubmüller, B Heymann, and P Tavan. ***Science***, 1996

11 OF Lange, N-A Lakomek, C Farès, ..., H Grubmüller, C Griesinger, and BL de Groot. ***Science***, 2008.

12 JJ Sieber, KI Willig, C Kutzner, ..., SW Hell, H Grubmüller, and T Lang. ***Science***, 2007

13 S Takamori, M Holt, ..., BL de Groot, J Klingauf, H Grubmüller, J Heuser, F Wieland, and R Jahn. ***Cell***, 2006

Download ➔ poster and publications PDFs