

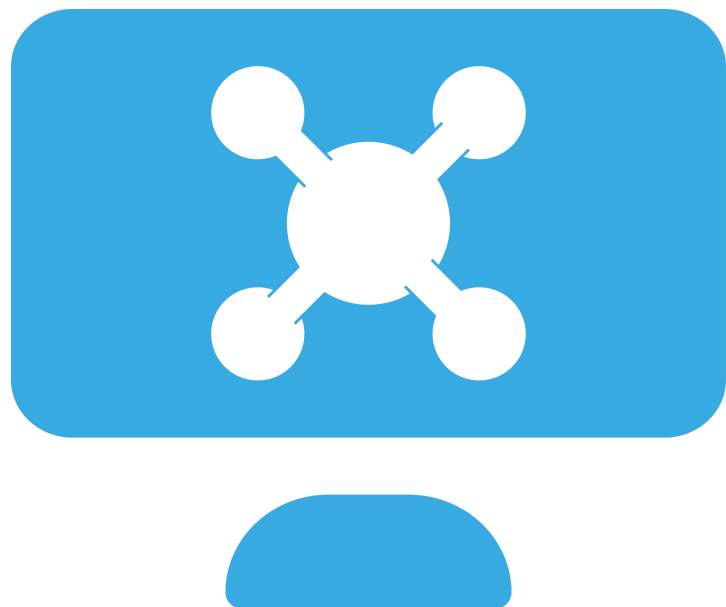
Molecular dynamics simulation for exascale supercomputing era: scientific research and software engineering challenges

Vedran Miletić 😎, *Matea Turalija* 😎

😎 Group for Applications and Services on Exascale Research Infrastructure (GASERI), Faculty of Informatics and Digital Technologies (FIDIT), University of Rijeka

Invited lecture at the **Computational Chemistry Day** in Zagreb, Croatia, 16th of September, 2023.

First off, thanks to the organizers for the invitation

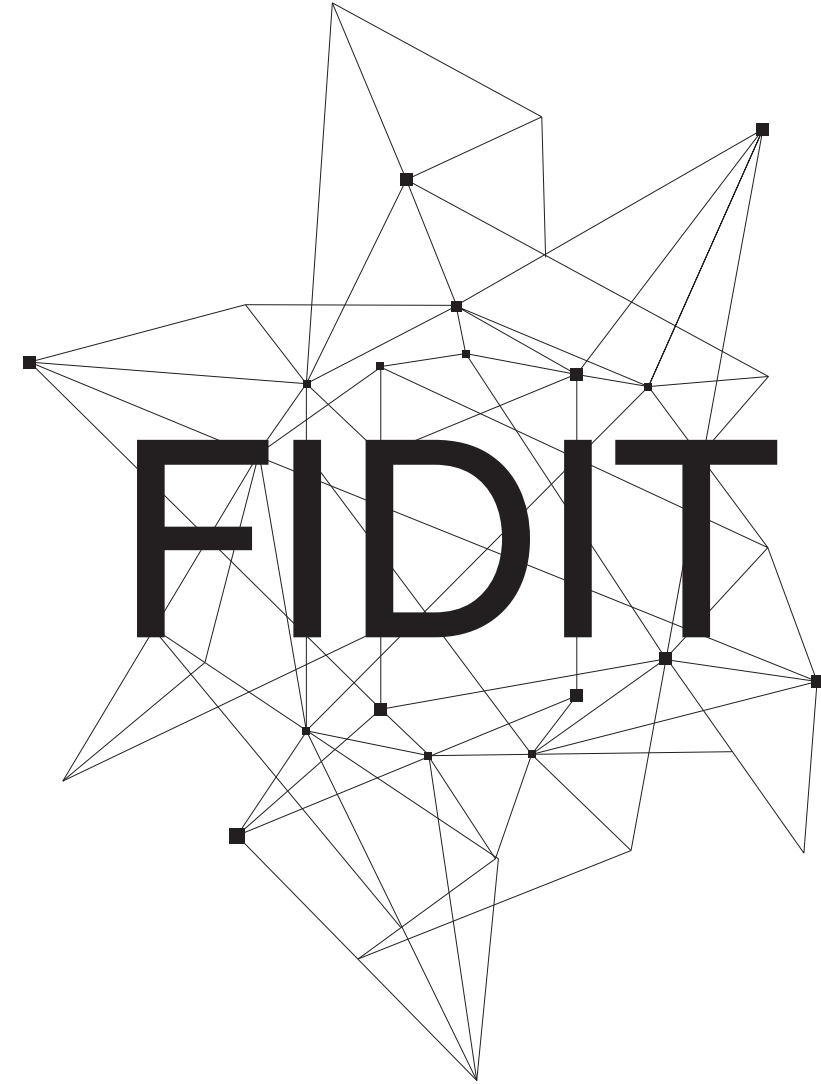


**COMPUTATIONAL
CHEMISTRY
DAY 2023**

Image source: [Computational Chemistry Day Logo](#)

Faculty of Informatics and Digital Technologies (FIDIT)

- founded in 2008. as *Department of Informatics* at University of Rijeka
- became *Faculty of Informatics and Digital Technologies* in 2022.
- **several laboratories** and research groups; scientific interests:
 - computer vision, pattern recognition
 - natural language processing, machine translation
 - e-learning, digital transformation
 - **scientific software parallelization, high performance computing**



Group for Applications and Services on Exascale Research Infrastructure (GASERI)

- focus: optimization of computational biochemistry applications for running on modern exascale supercomputers
- vision: high performance algorithms readily available
 - for academic and industrial use
 - in state-of-the-art open-source software
- members: [Dr. Vedran Miletić](#) (PI), [Matea Turalija](#) (PhD student), [Milan Petrović](#) (teaching assistant)



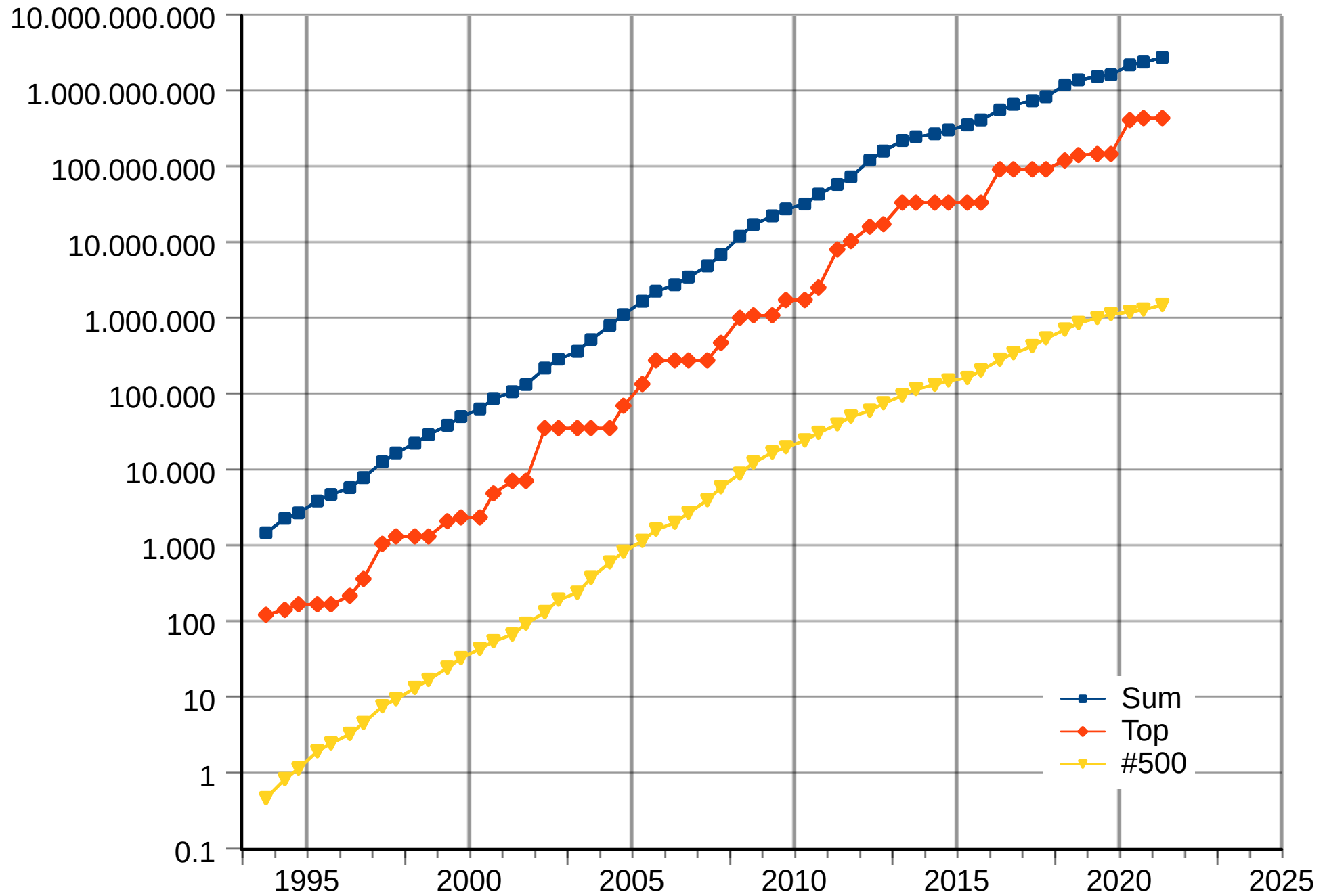
GASERI

How did we end up here, in computational chemistry?

- Prof. Dr. [Branko Mikac](#)'s optical network simulation group at FER, Department of Telecommunications
- What to do after finishing the Ph.D. thesis? 🤔
 - [NVIDIA CUDA Teaching Center](#) (later: GPU Education Center)
 - research in late [Prof. Željko Svedružić](#)'s Biomolecular Structure and Function Group and Group (BioSFGGroup)
- postdoc in [Prof. Frauke Gräter](#)'s Molecular Biomechanics (MBM) group at Heidelberg Institute for Theoretical Studies (HITS)
 - collaboration with GROMACS molecular dynamics simulation software developers from KTH, MPI-NAT, UMich, CU, and others

High performance computing

- a **supercomputer** is a computer with a high level of performance as compared to a general-purpose computer
 - also called high performance computer (HPC)
- measure: **floating-point operations per second** (FLOPS)
 - PC -> teraFLOPS; Bura -> 100 teraFLOPS, Supek -> 1,2 petaFLOPS
 - modern HPC -> 1 do 10 petaFLOPS, top 1194 petaFLOPS (~1,2 exaFLOPS)
 - future HPC -> 2+ exaFLOPS
- nearly exponential growth of FLOPS over time (source: [Wikimedia Commons File:Supercomputers-history.svg](#))



Languages used in scientific software development are also evolving to utilize the new hardware

- **C++**: C++11 (C++0x), C++14, C++17, C++20, C++23
 - Parallel standard library
 - Task blocks
- **Fortran**: Fortran 2008, Fortran 2018 (2015), Fortran 2023
 - Coarray Fortran
 - `DO CONCURRENT` for loops
 - `CONTIGUOUS` storage layout
- New languages: Python, Julia, Rust, ...

More heterogeneous architectures require complex programming models

- different types of accelerators
 - GPUs (half, single, double precision), TPUs/TCGPUs, FPGAs
 - in-network and in-storage computation (e.g. [BlueField DPU](#))
- several projects to adjust existing software for the exascale era
 - [Software for Exascale Computing](#) (SPPEXA)
 - [Exascale Computing Project](#) (ECP)
 - [European High-Performance Computing Joint Undertaking](#) (EuropHPC JU)

SPPEXA project GROMEX

- full project title: Unified Long-range Electrostatics and Dynamic Protonation for Realistic Biomolecular Simulations on the Exascale
- principal investigators:
 - [Helmut Grubmüller](#) (Max Planck Institute for Biophysical Chemistry, now Multidisciplinary Sciences, MPI-NAT)
 - [Holger Dachsel](#) (Jülich Supercomputing Centre, JSC)
 - [Berk Hess](#) (Stockholm University, SU)
- building on previous work on molecular dynamics simulation performance evaluation: [Best bang for your buck: GPU nodes for GROMACS biomolecular simulations](#) and [More bang for your buck: Improved use of GPU nodes for GROMACS 2018](#)

Parallelization for a large HPC requires domain decomposition

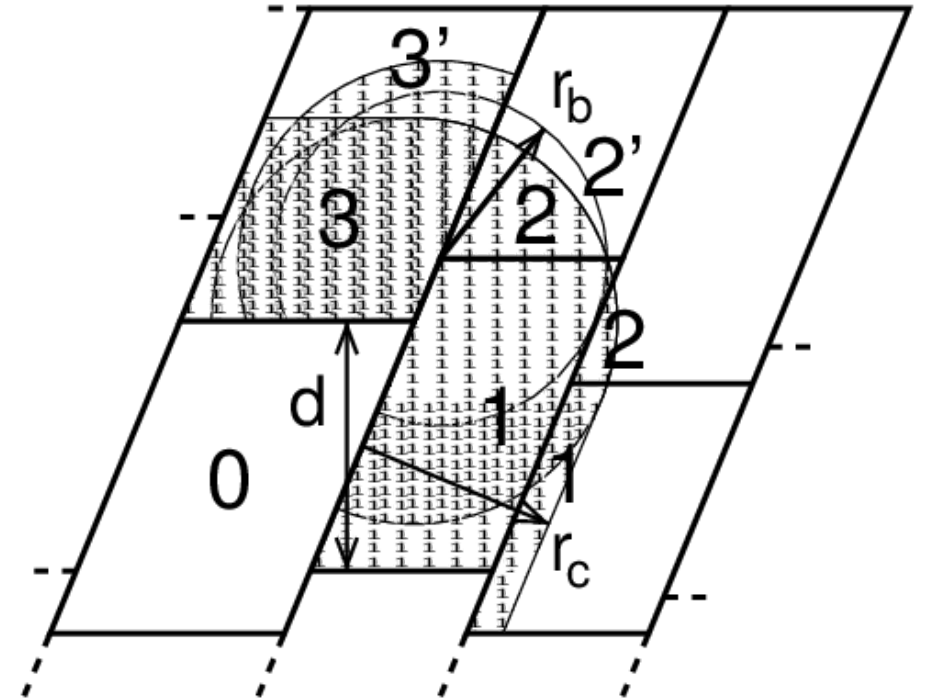
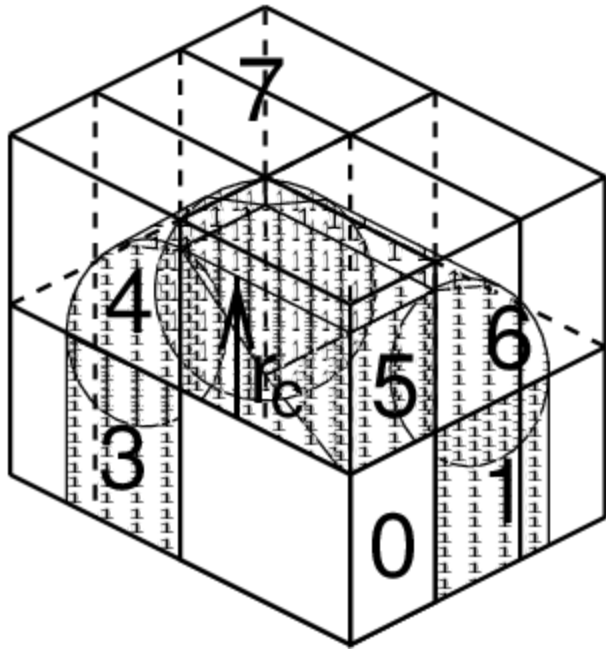
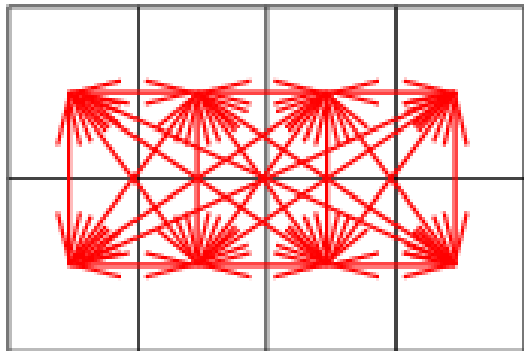


Image source for both figures:
[Parallelization \(GROMACS Manual\)](#)

Domain decomposition and particle mesh Ewald

8 PP/PME ranks



6 PP ranks

2 PME ranks

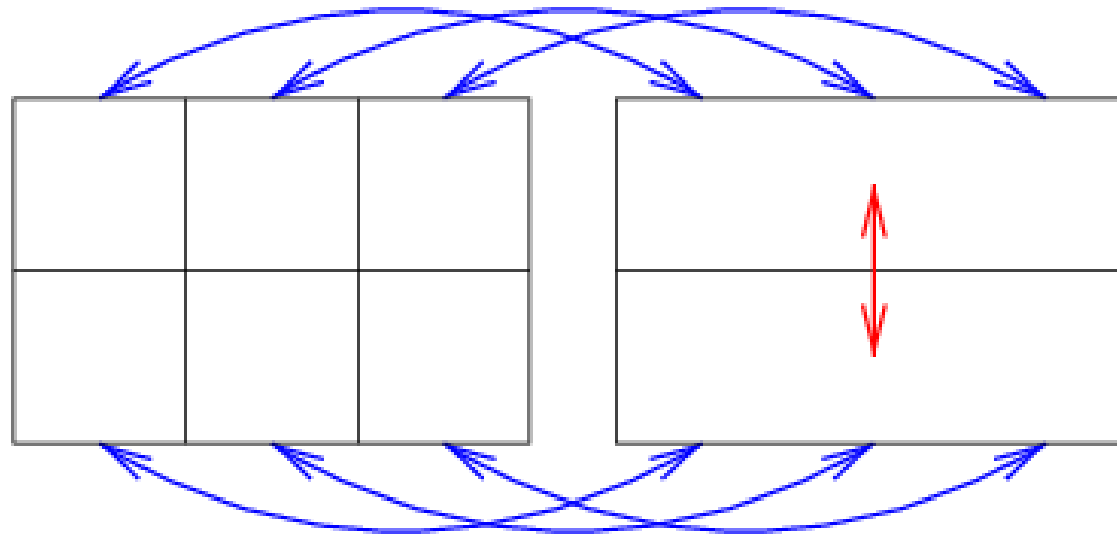


Image source for both figures: [Parallelization \(GROMACS Manual\)](#)

This approach still does not scale indefinitely. Is there an alternative?

GROMEX, part of SPPEXA

The particle mesh Ewald method (PME, currently state of the art in molecular simulation) does not scale to large core counts as it suffers from a communication bottleneck, and does not treat titratable sites efficiently.

The fast multipole method (FMM) will enable an efficient calculation of long-range interactions on massively parallel exascale computers, including alternative charge distributions representing various forms of titratable sites.

[SPPEXA Projects - Phase 2 \(2016 - 2018\)](#)

FlexFMM



Image source: [Max Planck Institute for Multidisciplinary Sciences FlexFMM](#)

- continuation of SPPEXA (2022 - 2025)
- project partners:
 - JSC: [Computational time optimized exascale simulations for biomedical applications](#)
 - MPI-NAT: [FMM-based electrostatics for biomolecular simulations at constant pH](#)
- our group is collaborating on the project via our colleagues at MPI-NAT

Our GROMACS developments: generalized FMM

- molecular dynamics simulations are periodic with various [simulation box types](#): cubic, rhombic dodecahedron; present design and implementation of the fast multipole method supports *only cubic* boxes
 - many useful applications (materials, interfaces) fit well into rectangular cuboid boxes, not cubic -> **Matea's PhD thesis research**

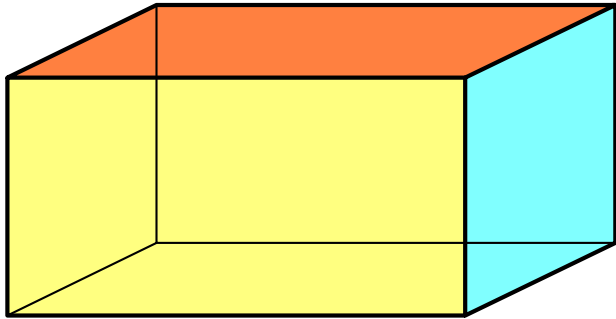


Image source: [Wikimedia Commons File:Cuboid no label.svg](#)

- it is possible to also support rhombic dodecahedron: ~30% less volume => ~30% less computation time per step required

Our GROMACS developments: NVIDIA BlueField



Image source: [Wikimedia Commons File:NVIDIA logo.svg](#)

- funded by NVIDIA, inspired by custom-silicon Anton 2 supercomputer's hardware and software architecture
- heterogeneous parallelism presently uses NVIDIA/AMD/Intel GPUs with CUDA/SYCL, also use NVIDIA BlueField DPUs with DOCA
- first publication came out last year: Turalija, M., Petrović, M. & Kovačić, B. [Towards General-Purpose Long-Timescale Molecular Dynamics Simulation on Exascale Supercomputers with Data Processing Units](#)
- DOCA 2.0 improved RDMA support, which eases our efforts

Our GROMACS developments: weight factor expressions and generalized flow simulation (1/4)

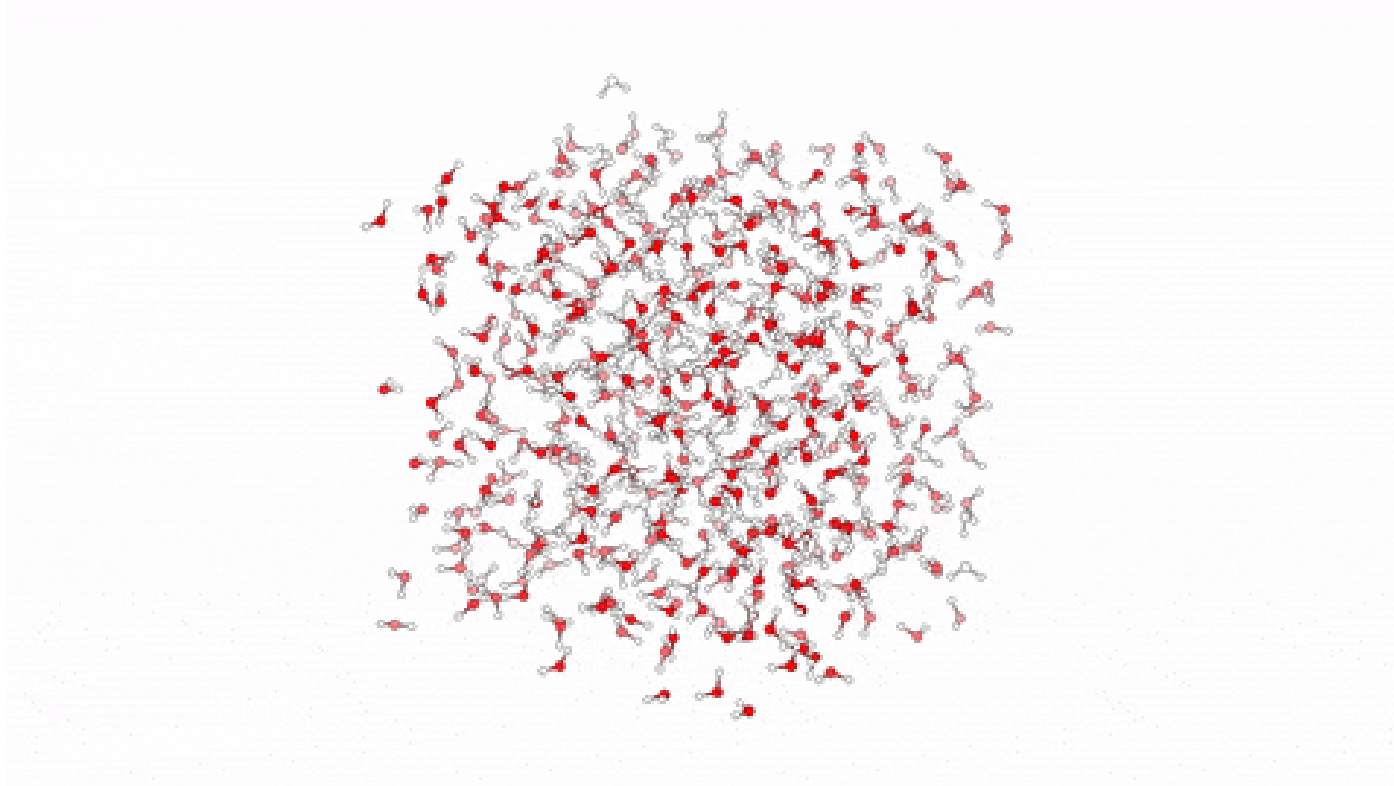


Image source: [Wikimedia Commons File:MD water.gif](#)

Our GROMACS developments: weight factor expressions and generalized flow simulation (2/4)

- Flow is a movement of solvent atoms
- Pulling all solvent atoms works when no other molecules except water are present in the simulation
- Just a slice of solvent molecules should be pulled to allow the solvent atoms to interact with the biomolecule(s) without being "dragged away"

Our GROMACS developments: weight factor expressions and generalized flow simulation (3/4)


 A rectangular water box with ~32,000 water molecules is shown.

Image source: Biophys. J. 116(6), 621–632 (2019). [doi:10.1016/j.bpj.2018.12.025](https://doi.org/10.1016/j.bpj.2018.12.025)

Our GROMACS developments: weight factor expressions and generalized flow simulation (4/4)

- Atom weight = dynamic weight factor computed from the expression **x** weight factor specified in the parameters file **x** atom mass-derived weight factor
- Dynamic weight factor (and atom weight) recomputed in each simulation step
- Weight factor expression variables:
 - Atom position in 3D (`x` , `y` , `z`)
 - Atom velocity in 3D (`vx` , `vy` , `vz`)
- Examples:
 - Atom weight factor is sum of squares of positions: `x^2 + y^2 + z^2`
 - Atom weight factor is a linear combination of velocities: `1.75 * vx + 1.5 * vy + 1.25 * vz`

Our *potential* GROMACS developments

- Monte Carlo ([Davide Mercadante](#), University of Auckland)
 - many efforts over the years, none with broad acceptance
 - should be rethought, and then designed and implemented from scratch with exascale in mind
- polarizable simulations using the classical Drude oscillator model ([Justin Lemkul](#), Virginia Tech)
 - should be parallelized for multi-node execution
- other drug design tools such as Random Acceleration Molecular Dynamics ([Rebecca Wade](#), Heidelberg Institute for Theoretical Studies and Daria Kokh, Cancer Registry of Baden-Württemberg)

Affiliation changes (1/2)

- starting October, Matea will be joining [Faculty of Medicine, University of Rijeka](#)
- PhD topic staying as planned and presented here

Image source: [Wikimedia Commons File:Medicinski fakultet Rijeka 0710 1.jpg](#)



Affiliation changes (2/2)

- several days ago, I joined [Max Planck Computing and Data Facility](#) in Garching near Munich, Germany as a part of [HPC Application Support Division](#)
- focus areas:
 - improving functionality and performance of lambda dynamics (free energy calculations)
 - developing fast multipole method implementation

Image source: [Wikimedia Commons](#)

[File:110716031-TUM.JPG](#)



**Thank you for your
attention**



GASERI website: group.miletic.net

