Molecular Dynamics simulations with GROMACS: GPU Acceleration

GROMACS is the among the most widely used softwares for Molecular Dynamics calculations. These calculations make use of Cloud GPU acceleration, facilitating biomolecular research.

seeweb
THINK CLOUD

Introduction to Molecular Dynamics

In 1648 Isaac Newton postulated the famous three laws that bear his name, from which classical physics-mathematics began [1].

• ***Lex. I.*** *Corpus omne perseverare in statu suo quiescendi vel movendi uniformiter in directum, nili quatenus a viribus impressis cogitur statum illum mutare.*

• ***Lex. II.*** *Mutationem motus proportionalem esse vi motrici impressae, et fieri secundum lineam rectam qua vis illa imprimitur.*

• ***Lex. III.*** *Actioni contrariam semper et aequalem esse reactionem: sive corporum duorum actiones in se mutuo semper esse aequales et in partes contrarias dirigi.*

Isaac Newton. *Philosophiae Naturalis Principia Mathematica. London, 1686.*

Three hundred and seventy-five years after the publication of what is considered one of the greatest scientific masterpieces, the *Principia Mathematica*, the same laws enunciated to describe the motion of the stars and planets still find application in simplifying the description of the atomic world through Molecular Dynamics [2].

Development and Applications of Molecular Dynamics

Molecular dynamics (MD) is a computer simulation method for analyzing the physical movements of atoms and molecules. The atoms and molecules are allowed to interact for a fixed period of time, giving a view of the dynamic "evolution" of the system under investigation.

The first MD simulations were carried out by Alder and Wainwright [3] in the 1950s, at the dawn of the computer age, to study simple fluids by models representing atoms as rigid disks and spheres. The simulated system consisted of just 32 atoms.

The **first simulation of a biological system**, specifically an enzyme, dates back to the late 1970s and was published by McCammon et al. in the prestigious journal Nature [4].

The study of the theoretical foundations that allowed such results earned the Nobel Prize in Chemistry in 2013 to M. Karplus, M. Levitt and A. Warshel [5,6], who can be considered the fathers of modern MD techniques. Today, 70 years after Alder's historic publication, it is possible **to simulate systems represented by several hundred thousand atoms**, obtaining data within days or weeks, thanks to the use of highly specialized hardware and software. Very recently, some research groups have been able to simulate entire viral particles [7,8] or even entire living cells [9] characterized by increasing complexity [10].

Progress and dissemination of MD in scientific research

MD is among the most time-consuming computational techniques because it is very iterative in nature and calculus-based. Since calculations are performed billions of times, this means that **millions of lines of code are involved**.

In recent decades, the use of MD has become increasingly accessible and widespread in scientific research, until it has become the **method of choice for the theoretical study of the atomic detail of biomolecules**.

In 1997, the articles using MD published in the most influential scientific journals were fewer than 175. After twenty years, in 2017, this number exceeded 1,000 [11]. Undoubtedly, this progress has been strongly driven by the development of new hardware (thanks to the use of graphics processing units, GPUs), software and algorithms (parallel programming techniques and more efficient MD codes), but also in the design of increasingly accurate physical models ('force-fields').

Accelerating MD simulations with GPU.

A graphics processing unit (GPU) is a highly specialized electronic circuit initially designed to accelerate computer graphics and image processing. After their initial design and due to their **"parallel" structure**, GPUs were found to be useful for non-graphic calculations aimed at solving computational problems divided into small, simple, independent tasks that can be performed simultaneously (with little or no communication between them) just as in the case of DM simulations. Therefore, the advent of **GPU computing has brought a real revolution in DM**, making accessible and relatively inexpensive the high performance that was previously reserved for supercomputers.

As an example, a GPU-equipped computer can run simulations on average up to ten times faster than the same CPU-equipped computer. Moreover, to achieve the same performance, the same CPU-equipped computer would have to be five times larger than a GPU-equipped computer: consequently, using GPUs also halves power consumption. To date, **most softwares for MD calculations are GPU compatible** [12].

Among them, GROMACS undoubtedly represents one of the most common and efficient choices.

The most recent version of GROMACS (2023), allows reducing the overhead between GPU and CPU, greatly increasing the performance of biomolecular simulation software [13]

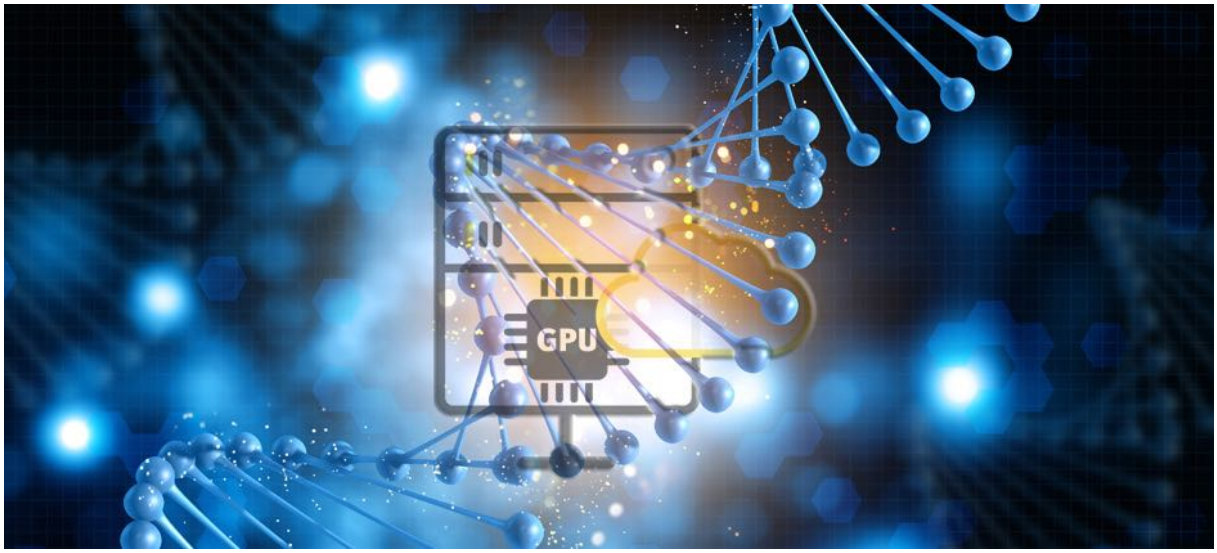
GROMACS, GPU and Cloud Computing

"GROMACS is one of the world's most widely used open source molecular dynamics applications, and it's easy to see why. The simulations we can conduct with the application grants us better understanding into things as small as the proteins in our bodies to as large as the galaxies in the universe. Most notably, our work with GROMACS – developed and optimized with oneAPI – allows Intel to have a hand in significant advances in drug discovery and expands GROMACS open development across multiple compute architectures. And this is all while collaborating with the open source community that we so greatly value."

– ***Roland Schulz**, parallel software engineer at Intel*

GROMACS (*GROningen MAChine for Simulations*) is an **open-source software designed for high-performance atomistic simulations** of proteins, lipids, carbohydrates, nucleic acids and particles. Initially developed by the University of Groningen, the project is supported by other universities and the European community through the Horizon2020 program. It is used primarily for MD simulations of biological systems (although it can be used for simulations of other types of systems) and offers a comprehensive set of various types of computation, input preparation and output analysis.

GROMACS supports all the most common algorithms expected from a modern MD software, but includes some innovative and distinctive features that distinguish it from its competitors [14] :



• GROMACS can make simultaneous use of both CPU and GPU available in a system. There are options to statically and dynamically balance the load between the different resources.

• GROMACS is user-friendly, with topologies and parameter files written in clear text format. There is a lot of consistency checking, and clear error messages are issued when something is wrong.

• There is no scripting required - all programs use a simple interface with command line options for input and output files. You can always get help on the options by using the -h option, or use the extensive manuals provided free of charge in electronic or paper format. There is ongoing work on a Python API which enables scripting of simulation setup, running and analysis.

• Both run input files and trajectories are independent of hardware endian-ness, and can thus be read by any version GROMACS.

• GROMACS can write output using lossy compression, which provides a very compact way of storing trajectory data. The accuracy can be selected by the user.

• GROMACS comes with a large selection of flexible tools for trajectory analysis and the output formats are also supported by all major analysis and visualisation packages.

• GROMACS can be run in parallel, using either the standard MPI communication protocol, or via our own "Thread MPI" library for single-node workstations.

• GROMACS contains several state-of-the-art algorithms that make it possible to extend the time steps is simulations significantly, and thereby further enhance performance without sacrificing accuracy or detail.

• The package includes a fully automated topology builder for proteins, sugars, lipids and nucleic acides, even multimeric structures.

• GROMACS operates within a multilevel parallelization scheme that distributes computational work among sets of simulations.

• GROMACS allows systems to be described with triclinic unit cells; therefore, complex geometries such as rhombic dodecahedron, truncated octahedron, or hexagonal cells are supported.

The key to optimizing the performance of MD simulations lies in understanding the **use of hardware by DM algorithms**. In particular, GROMACS combines several parallelization techniques, including MPI and OpenMP, GPU "offloading," and separable calculations of long-range electrostatic interactions.

GROMACS can performs several types of computationally intensive calculations on the GPU. In detail, one of the biggest performance benefits is the computation of short-range nonbonding interactions (i.e.: Coulomb and van der Waals) on the GPU. The ability to perform these types of calculations in parallel (in which each MPI rank takes advantage of a GPU to calculate the short-range interactions) further optimizes performance. In contrast, the computation of long-range interactions could not be distributed across multiple GPUs until early 2023. Following the release of the GROMACS 2023 version, the performance of MD simulations could be significantly improved, making the system even more scalable than it was previously [13,14].

GROMACS is used in healthcare by biotechnology organizations, universities and research centers, education, pharmaceutical organizations, hospitals and clinics.

However, despite the significant and ongoing technological advances in hardware and software that we have just highlighted, running **MD simulations remains a very computationally intensive task**.

Where do researchers find the necessary computational time? Established solutions include computing centers at universities and research institutes, national supercomputing centers, local clusters, and single-node workstations, each with advantages and disadvantages regarding ease of access to resources, their availability and relative cost.

Recently, a viable solution to these issues is cloud computing, given the ability to better manage computation resources: that is, in this way the user has access to high performance computing (HPC) systems that can be easily adapted and reshaped according to specific needs at a given time. As a result, cloud computing resources can be finely tuned with the aim of optimizing costs or minimizing computing time.

Several studies have shown that GROMACS performance and P/C ratio (performance/cost, i.e., the amount of data produced per € invested) are highly dependent on the available hardware [16]. In this regard, single nodes workstations equipped with multiple GPUs provide better performance. On the other hand, the best performing GPUs to date are priced in the range of several (sometimes tens) thousands of euros and, among other hardware components, are those most subject to economic fluctuations.

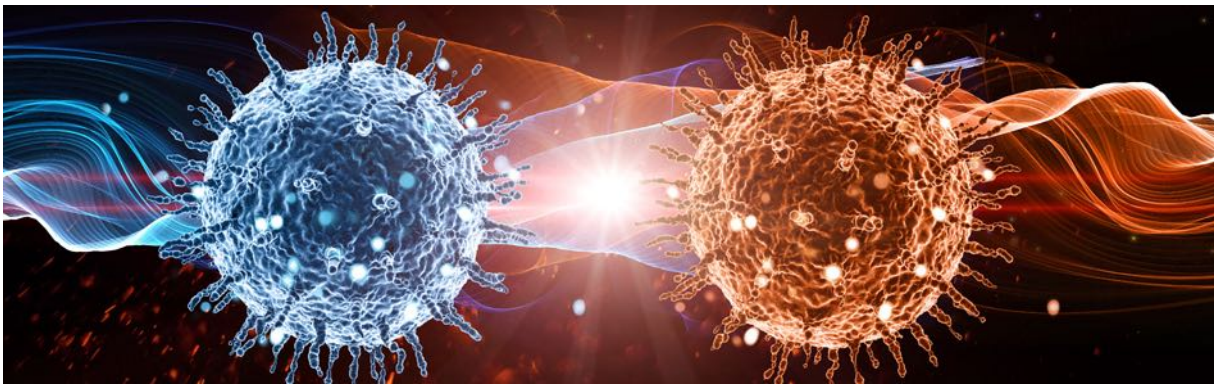
Again, **cloud computing proves to be an excellent choice that simultaneously maximizes performance and P/C ratio**. This approach provides the user with greater ease of use of the system and a drastic reduction in the time required to install and test softwares – mandatory and preliminary operations to MD calculations – that can be tedious even for experienced users.

GROMACS in Biomedical Research: some application examples

MD simulations are constantly and successfully applied at different stages of research and development of new bioactive molecules, against a wide variety of diseases, to obtain key information about the three-dimensional structure of biomolecules and their dynamics. This provides insight into the structure-function relationship that can guide the **drug discovery and design process**.

A fascinating example of the use of MD simulations is contextualized in the pandemic of COVID-19, which is being fought not only in hospitals and laboratories, but also in large computing centers where the very powerful supercomputers are housed.

Recently, it has been possible through MD to simulate and predict the behavior of SARS-CoV-2 coronavirus particles in countless contexts and conditions. In one of the most advanced studies in the field [7], the research group led by Dr. R. Amaro, simulated a coronavirus viral particle within an aerosol droplet at the atomic scale – a system composed of more than a billion atoms – tracing their movements with precision on the order of millionths of a second. This has been made possible by software such as GROMACS and has been instrumental to studying interactions and possible effects on contagiousness.



Even more recently, the same research group carried out a similar study, but this time investigating the structural and dynamic properties of influenza virus glycoproteins [8]. The surface glycoproteins of the virus are responsible for infection and, as they mutate very rapidly, make it difficult to develop a universal vaccine. This study has provided insight into previously unknown details about the dynamics of these important glycoproteins and their interaction with each other, suggesting new strategies for the development of influenza vaccines and antiviral drugs.

Another example of the increasingly close interconnection between laboratory and simulation activities in the biomedical field comes from a scientific research published last May in the journal Nature [15]. This work allowed us to understand the three-dimensional structure of a human olfactory receptor, called OR51E2. This is an important step, allowing us to understand more about the mechanism of one of our most complex and least understood senses (the human sense of smell involves about 400 different receptors!). Each of the hundreds of thousands of scents we can perceive is made up of a mixture of different

odor molecules, each of which, in turn, can be detected by certain types of receptors: it is, therefore, a truly huge puzzle which the brain must solve every time the nose recognizes something in the air. The biggest obstacle toward understanding this mechanism at the molecular level, however, is the considerable difficulty in obtaining information about the structure of olfactory receptors in the laboratory, a feat considered nearly impossible. In this study, the authors were able to greatly improve the sensitivity of the instruments used, which allowed them to work with even the smallest amounts of receptors available.

"*This is the first 'picture' of an olfactory molecule interacting with one of our olfactory receptors*," explains Aashish Manglik, one of the authors.

In this case, MD simulations were used to investigate details about **structural atomic changes that experiments cannot achieve**: Manglik et al. were able to understand the changes in receptor structure and the specific molecular interactions that occur when it binds the odor molecule to initiate the signaling process underlying the sense of smell.

In the field of **medicinal chemistry**, an approach called *fragment-based drug design* (FBDD) has become increasingly popular and is now recognized as a viable alternative to HTS (High Throughput Screening) procedures. It consists in examining a very small initial number (on the order of a few thousand) of small molecules (*fragments*), which are then processed until new lead compounds are identified for development into drugs.

Currently, the **FBDD approach** has led to the development of six clinically approved drugs. The small size of the chemical fragments used, however, results in very weak and transient binding to the drug target. This makes it very complicated to determine, through classical experiments, the three-dimensional structure assumed by the fragments following binding to the target protein.

Conclusions

Once again, MD simulations represent a method for overcoming the limitation described above. Thanks to the *Anton 2* supercomputer, it was possible to perform MD simulations on the order of milliseconds – among the longest ever performed – which allowed the **determination of binding sites and modes of some chemical fragments** towards the PTP1b protein, an important – yet notoriously difficult to study – pharmaceutical target [17]. Subsequent experiments proved the validity of the predictions made by MD. This recent study represents the first demonstration that MD simulations can be used *prospectively* (not retrospectively; a well-established use of this computational technique) to identify unknown binding sites on the surface of target. As the authors themselves point out, such an approach is not easily applicable on a large scale (i.e., screening libraries of hundreds of fragments), although the use of **specialized hardware** may make it possible.

Nevertheless, with increasing computational power and the development of increasingly accurate physical models, the use of MD simulations on these time scales will also become increasingly common in the context of identifying chemical fragments of pharmacological interest, complementing existing technologies applied in the early stages of drug development.

In this context, the acceleration provided by GPU computing services, but also their ability to optimize technology investments by improving the outcome, represents a key strategy in ‘speeding up the timeline of scientific research.

Bibliography

1. Isaac Newton. **Philosophiae Naturalis Principia Mathematica**. London, 1686.
2. Danilo Roccatano. **Introduzione alla Dinamica Molecolare**. Maggio 2015
3. Alder BJ, and Wainwright TE (1957). **Phase Transition for a Hard Sphere System**. J Chem Phys 27, 1208–1209.
4. McCammon JA, Gelin BR, and Karplus M (1977). **Dynamics of Folded Proteins**. *Nature* 267, 585–590.
5. Levitt M, and Lifson S (1969). **Refinement of Protein Conformations Using a Macromolecular Energy Minimization Procedure**. *Journal of Molecular Biology* 46, 269–279.
6. Lifson S, and Warshel A (1968). **Consistent Force Field for Calculations of Conformations Vibrational Spectra and Enthalpies of Cycloalkane and N-Alkane Molecules**. *J Chem Phys* 49, 5116–5129.
7. Jacob D. Durrant, Sarah E. Kochanek, Lorenzo Casalino, Pek U. Ieong, Abigail C. Dommer, and Rommie E. Amaro. **Mesoscale All-Atom Influenza Virus Simulations Suggest New Substrate Binding Mechanism**. ACS Central Science 2020 6 (2), 189-196.
8. Dommer A, Casalino L, Kearns F, Rosenfeld M, Wauer N, Ahn SH, Russo J, Oliveira S, Morris C, Bogetti A, Trifan A, Brace A, Sztain T, Clyde A, Ma H, Chennubhotla C, Lee H, Turilli M, Khalid S, Tamayo-Mendoza T, Welborn M, Christensen A, Smith DG, Qiao Z, Sirumalla SK, O'Connor M, Manby F, Anandkumar A, Hardy D, Phillips J, Stern A, Romero J, Clark D, Dorrell M, Maiden T, Huang L, McCalpin J, Woods C, Gray A, Williams M, Barker B, Rajapaksha H, Pitts R, Gibbs T, Stone J, Zuckerman DM, Mulholland AJ, Miller T 3rd, Jha S, Ramanathan A, Chong L, Amaro RE. **#COVIDisAirborne: AI-enabled multiscale computational microscopy of delta SARS-CoV-2 in a respiratory aerosol**. Int J High Perform Comput Appl. 2023 Jan;37(1):28-44.
9. Thornburg ZR, Bianchi DM, Brier TA, et al. **Fundamental behaviors emerge from simulations of a living minimal cell.** Cell. 2022 Jan;185(2):345-360.e28.
10. Stevens JA, Grünewald F, van Tilburg PAM, König M, Gilbert BR, Brier TA, Thornburg ZR, Luthey-Schulten Z and Marrink SJ (2023) **Molecular dynamics simulation of an entire cell.** *Front. Chem.* 11:1106495.
11. Hollingsworth SA, Dror RO. **Molecular Dynamics Simulation for All. Neuron**. 2018 Sep 19;99(6): 1129-1143.
12. https://images.nvidia.com/content/tesla/pdf/Molecular-Dynamics-July-2017-MB-slides.pdf
13. https://developer.nvidia.com/blog/a-guide-to-cuda-graphs-in-gromacs-2023/
14. GROMACS 2023 User's Manual (https://doi.org/10.5281/zenodo.8134388)
15. Billesbølle, C.B., de March, C.A., van der Velden, W.J.C. *et al*. **Structural basis of odorant recognition by a human odorant receptor**. *Nature* 615, 742–749 (2023).
16. Kutzner C, Páll S, Fechner M, Esztermann A, de Groot BL, Grubmüller H. Best bang for your buck: **GPU nodes for GROMACS biomolecular simulations**. J Comput Chem. 2015 Oct 5;36(26):1990-2008.
17. Jack B. Greisman, Lindsay Willmore, Christine Y. Yeh, Fabrizio Giordanetto, Sahar Shahamadtar, Hunter Nisonoff, Paul Maragakis, and David E. Shaw. **Discovery and Validation of the Binding Poses of Allosteric Fragment Hits to Protein Tyrosine Phosphatase 1b: From Molecular Dynamics Simulations to X-ray Crystallography**. *Journal of Chemical Information and Modeling* 2023 63 (9), 2644-2650.

seeweb

THINK CLOUD

Via Armando Vona 66
03100, Frosinone

Via Caldera, 21
Blue Building ala 1
20153 Milano

[T] +390775880041
[@] info@seeweb.it
[W] www.seeweb.it

cloud server gpu

Samuele Di Cristofano, PhD
Laboratorio di Bioinformatica, Prof. D. Raimondo.
Dip.to di Medicina Molecolare.
Facoltà di Farmacia e Medicina - Sapienza Università di Roma.