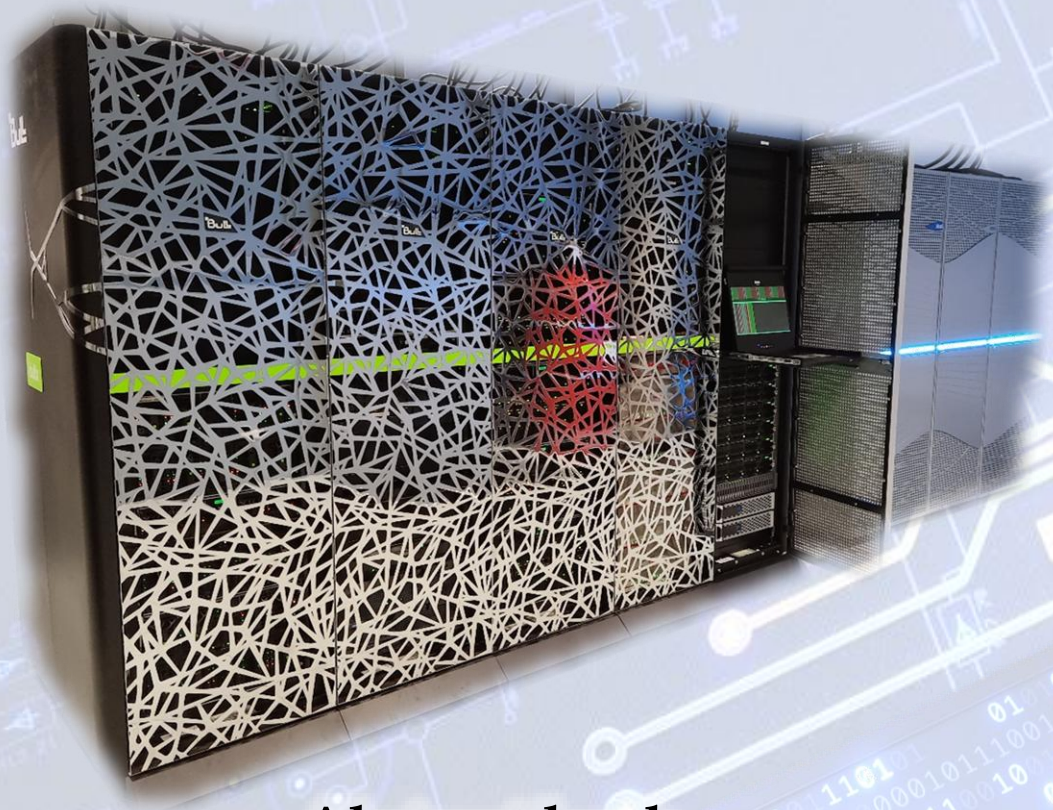
4TH EUROCC VILNIUS HACKATHON & WORKSHOP ON USING HPC



Abstract book

https://doi.org/10.5281/zenodo.15754592

https://www.eurocc-lithuania.lt/events_2025-06-27/

June 27, 2025

Vilnius, Lithuania

Hackathon & Workshop organizers

Local organizing committee

Mindaugas Mačernis
Laura Baliulytė

Scientific committee

Mindaugas Mačernis
Laura Baliulytė

Funding

Co-funded by the European Union

EuroHPC Joint Undertaking

Funded by the European Union. This work has received funding from the European High Performance Computing Joint Undertaking (JU) and Germany, Bulgaria, Austria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, Greece, Hungary, Ireland, Italy, Lithuania, Latvia, Poland, Portugal, Romania, Slovenia, Spain, Sweden, France, Netherlands, Belgium, Luxembourg, Slovakia, Norway, Türkiye, Republic of North Macedonia, Iceland, Montenegro, Serbia under grant agreement No 101101903.

Bendrai finansuoja Europos Sąjunga

Projektas bendrai finansuojamas 2021–2027 metų ES fondų investicijų programos (sutartis Nr. 10-051-P-0001).

EuroCC2-EuroCC4SEE
Project Organiser

Vilniaus universitetas 1579 Universitas Vilnensis

Project Implementers

Vilniaus universitetas 1579 Universitas Vilnensis — Fizikos fakultetas

Vilniaus universitetas 1579 Universitas Vilnensis — Matematikos ir informatikos fakultetas

ktu kauno technologijos universitetas 1922

VILNIUS TECH

Lietuvos hidrometeorologijos tarnyba

Molecular Dynamics Statistical Analysis: Probabilistic Modeling of Molecular Geometry from MD Data

Jonas Franukevičius[1], Mindaugas Mačernis[1]

[1]*Institute of Chemical Physics, Faculty of Physics, Vilnius University*

E-mail: jonas.franukevicius@ff.vu.lt

We introduce *Molecular Dynamics Statistical Analysis* (MDSA), an algorithm designed to statistically characterize the internal geometries of molecular systems from molecular dynamics (MD) simulations. MDSA provides a method for analyzing and representing the distribution of internuclear distances sampled from MD trajectories.

The algorithm operates in three stages. First, MDSA computes the uncorrelated probability densities for all pairwise internuclear distances, averaged over the MD simulation time. In the second stage, MDSA constructs a *Linear Combination of Gaussian Models* (LCGM) to approximate the joint distribution of the D-dimensional internuclear distance space. The LCGM comprises of preselected number of D-dimensional Gaussian distributions, where the amplitudes, means and standard deviations are optimized to best match the probability densities derived from the MD data and lastly, the inferred structures from Gaussian models are reformated to Euclidean geometries and then compared to MD data.

The optimization is guided by a cost function composed of two terms. The first term quantifies the fidelity of the fit through the integral of squared differences between the MD-derived probability density and that generated by the LCGM. The second term is a *realizability cost*, which penalizes Gaussian models that do not correspond to physically possible three-dimensional molecular structures. This realizability cost is computed by evaluating the Gram matrix corresponding to a given configuration of internuclear distances (as defined by the Gaussian model parameters). A valid configuration requires that the Gram matrix is positive semi-definite and has rank exactly three, ensuring that the distances define a realizable molecular geometry in Euclidean space [1].

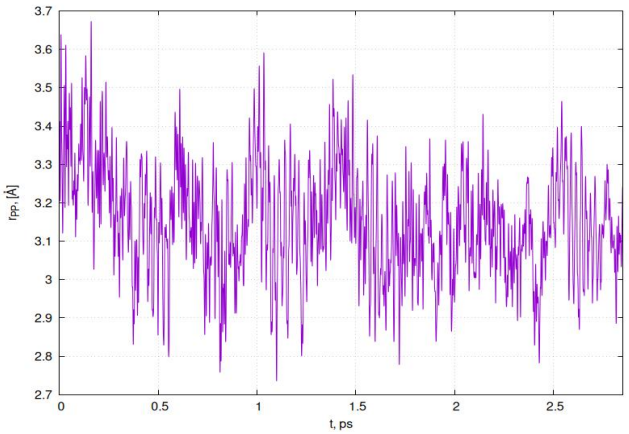


Fig. 1. P–P distance time dependence (violet) extracted from $CH(PPh_2)_2$ molecular dynamics data.

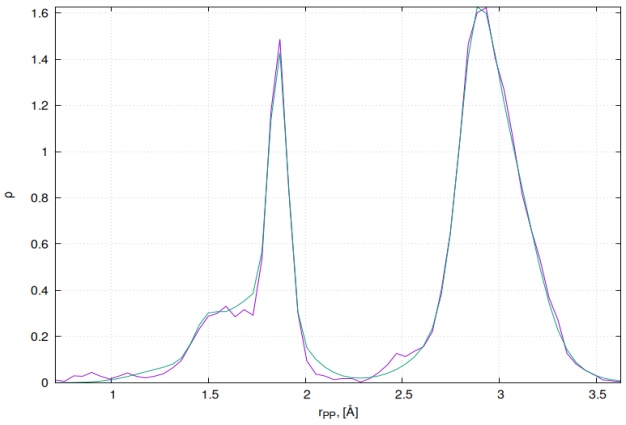


Fig. 2. Calculated probability density from MD data (violet) and its LCGM counterpart (green).

REFERENCES

[1] I. Dokmanić, R. Parhizkar, J. Ranieri and M. Vetterli, "Euclidean Distance Matrices Essential Theory, Algorithms and Applications", I*EEE Signal Processing Magazine vol. 32, no. 6, pp. 12-30* (2015).