# Machine learning approach to analytically represent potential energy surfaces: Application to He-CO2 bending potential energy surface

**Taha Selim and Alain Chancé**
MolKet SAS

December 28, 2024

This document details the development and application of a machine learning model to analytically represent the potential energy surface (PES) governing the bending mode of the He-CO2 van der Waals complex [5].

This potential energy surface has been published in the Journal of Chemical Physics, 159, 164310 (2023). 

Article's info:
DOI: `https://doi.org/10.1063/5.0174787`

This article includes results of the work done during the PhD study of Taha Selim at Institute of Molecules and Materials, Radboud University Nijmegen, the Netherlands [2]. For more info, please refer to Selim's PhD thesis:
`https://repository.ubn.ru.nl/handle/2066/311101`

Further information:
Molket SAS website: `https://www.molket.io`
Github repository: `https://github.com/molket-io/molket.jl`

# 1 Potential energy surface data

We start by outlining the data of the Fortran subroutine used to generate the potential energy surface (PES) of the He-$CO_2$ van der Waals complex is calculated using a Fortran subroutine that computes the interaction energy between the helium atom and the $CO_2$ molecule at four Jacobi coordinates.

The output is the interaction potential energy in atomic units (Hartree) as a function of the following four Jacobi coordinates: $V(Q, R, \theta, \phi)$.

The four Jacobi coordinates are defined as follows:

- **INPUT:**
  - `R0`: $CO_2$-He center of mass distance (in $a_0$). The distance between the helium atom and the center of mass of the $CO_2$ molecule.
  - `theta`: angle between the vector `R0` pointing from the $CO_2$ center of mass to the He nucleus and the z-axis (in radians)
  - `phi`: angle between the projection of the vector `R0` on the xy-plane and the x-axis (in radians)
  - `QQ`: dimensionless normal coordinate of $CO_2$ bend mode ($\nu_2$) [1]
    - See the paper for details and the description of the subroutine `Qtrans(Qcart,Qdimensionless)` below.
- **Output:**
  - `V`: interaction potential energy in atomic units (Hartree)

The subroutine is written in `Fortran` and it is called from `Julia` using our notebook `HeCO2_4Dpot_test.ipynb`.

Thus, the Fortran subroutine generated a single point of the potential energy surface for a given set of Jacobi coordinates.

There are many ways to represent the data in a machine learning model. Here, we give two examples:

- **Direct representation:** The four Jacobi coordinates are used as input features and the potential energy is the output. This representation yields a table of data points. Example:

| $R0a_0$ | $\theta$(degrees) | $\phi$(degrees) | $QQ$(dimensionless) | $V\,\mathrm{cm}^{-1}$ |
|---|---|---|---|---|
| 5.7 | 90.0 | 0.0 | 0.0 | $-46.47$ |
| 5.8 | 90.0 | 0.0 | 1.0 | $-44.696$ |

- **Pseudo quantum number (PQN) representation**: The four Jacobi coordinates are used and represented as a single variable (e.g. a pseudo quantum number) and the potential energy is the output. Example:

| PQN | $R0a_0$ | $\theta$(degrees) | $\phi$(degrees) | $QQ$(dimensionless) | $V\text{cm}^{-1}$ |
|---|---|---|---|---|---|
| 1 | 5.7 | 90.0 | 0.0 | 0.0 | $-46.47$ |
| 2 | 5.8 | 90.0 | 0.0 | 1.0 | $-44.696$ |

Moreover, there are symmetries in the data that can be exploited to reduce the number of data points. For example, the potential energy surface is symmetric with respect to the $\theta$ in which the interaction energy is invariant under the transformation $\theta \rightarrow -\theta$ in the range $[0, \pi]$. This is not the case with respect to other vibrational mode, asymmetric stretch mode, where the symmetry in $\theta$ only happens at the linear rigid structure of $CO_2$ [2, 3, 4, 5].

# References


[1] Purdue University. Vibrations of carbon dioxide. URL: `https://www.chem.purdue.edu/jmol/vibs/co2.html`.

[2] Taha Selim. Quantum dynamics of molecules in space: Theoretical studies and efficient computational methods for collision-induced rovibrational transition rates in molecules. *PhD Thesis*, 2024. URL: `https://repository.ubn.ru.nl/handle/2066/311101`.

[3] Taha Selim, Arthur Christianen, Ad van der Avoird, and Gerrit C. Groenenboom. Multi-channel distorted-wave born approximation for rovibrational transition rates in molecular collisions. *J. Chem. Phys.*, 155:034105, 2021. `doi:10.1063/5.0058576`.

[4] Taha Selim, Ad van der Avoird, and Gerrit C. Groenenboom. Efficient computational methods for rovibrational transition rates in molecular collisions. *J. Chem. Phys.*, 157:064105, 2022. `arXiv:2206.04470`, `doi:10.1063/5.0102224`.

[5] Taha Selim, Ad van der Avoird, and Gerrit C. Groenenboom. State-to-state rovibrational transition rates for $co_2$ in the bend mode in collisions with he atoms. *J. Chem. Phys.*, 159:164310, 2023. `arXiv:2309.03781`, `doi:10.1063/5.0174787`.