PLACEHOLDER

Joan Tibau Terma

**Resum** — La simulació de dinàmica molecular (DM) és una tècnica computacional que estudia sistemes moleculars, proporcionant informació sobre l'estructura, l’energia, la dinàmica i altres propietats. Té aplicacions en la indústria farmacèutica, biologia, enginyeria de materials, física i química. S'utilitzen mètodes com Verlet per a resoldre les equacions del moviment simultàniament a mètodes de mecànica quàntica per a calcular les forces i energies d'interacció entre àtoms. Amb sistemes complexos, es requereixen tècniques d'optimització i paral·lelització. Les noves tècniques d'aprenentatge computacional, com les xarxes neuronals, ofereixen alternatives per abordar els problemes de complexitat de la DM. Aquest projecte té com a objectiu de l’estudi i l'aplicació de xarxes neuronals a simulacions físiques per a la predicció de propietats moleculars. S'utilitzat la toolbox SchNetPack2 que ptoporciona eines per a aplicar xarxes neuronals per a la predicció de propietats de bases de dades com ara QM9. S’ha realitzat una anàlisi detallada dels resultats obtinguts i s'han extret conclusions rellevants per a la millora del model.

**Paraules clau** — Analítica de Dades, xarxes neuronals, dinàmica molecular, base de dades QM9, SchNet, predicció de propietats, anàlisi de resultats, optimització.

**Abstract** — Molecular dynamics (MD) simulation is a computational technique that studies molecular systems, providing information about structure, energy, dynamics, and other properties. It has applications in the pharmaceutical industry, biology, materials engineering, physics, and chemistry. Numerical methods like Verlet or Gear are used to solve the equations of motion and calculate interaction forces. With complex systems, optimization and parallelization techniques are required. New computational learning techniques, such as neural networks, offer alternatives to address the complexity of MD. This project aims to study and apply neural networks to physical simulations for the prediction of molecular properties. The SchNetPack2 toolbox is used, providing tools for applying neural networks to predict properties from databases like QM9. A detailed analysis of the obtained results has been conducted, drawing relevant conclusions for model improvement.

**Index Terms** — Data Science, neural networks, molecular dynamics, QM9 database, SchNet, property prediction, result analysis, optimization.

—————————— ◆ ——————————

# 1 Introduccio:

L

ES simulacions de dinàmica molecular son una tècnica computacional avançada que permet l'estudi detallat de sistemes moleculars. Aquesta metodologia proporciona informació valuosa sobre l'estructura, l'energia, la dinàmica i altres propietats dels sistemes en qüestió. Les aplicacions de la dinàmica molecular són àmplies i diverses, abarcant àmbits com la indústria farmacèutica, la biologia, l'enginyeria de materials, la física i la química.

L'objectiu principal d'aquest treball és explorar i aplicar tècniques de xarxes neuronals en simulacions físiques, especialment en el camp de a la dinàmica molecular. Amb el desenvolupament i l'avanç de les xarxes neuronals, s'obre la porta a noves possibilitats per afrontar els desafiaments i les complexitats associades amb els calcus necesaris per obtener aquestes propietats.

El document està organitzat de la següent manera: en primer lloc, es proporciona una revisió l’state of the art del camp de les xarxes neuronals aplicades a dinámica molecular, tant bases de dades com models frameworks o toolbox existents. Seguidament es presenta l'eina SchNetPack2 i com s'ha utilitzat en aquest estudi. A continuacio, s'exposen els resultats obtinguts de les simulacions i s'analitzen en detall. Finalment, es presenten les conclusions rellevants i les possibles direccions futures per millorar i ampliar aquest treball.

————————————————

1. E-mail de contacte: jotite19@gmail.com
2. Menció realitzada: Computació
3. Treball tutoritzat per: Ramon Baldrich
4. Tutoria externa per: Jordi Faraudo
5. Curs 2022/23

A través d'aquesta investigació, es pretén contribuir al camp de la dinàmica molecular i les seves aplicacions mitjançant l'ús de tècniques d'aprenentatge computacional. S'espera que els resultats obtinguts proporcionin una millor comprensió de la utilitzacio de xarxes neuronals en el camp.

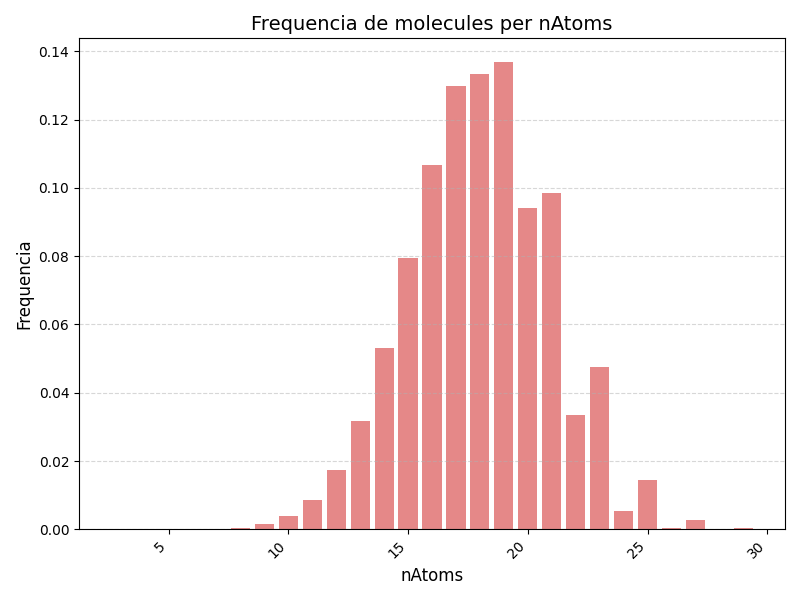
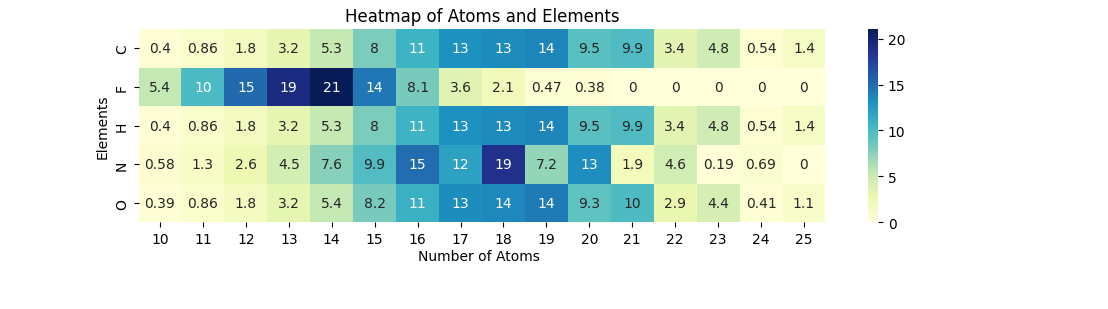
# 2 State of the Art:

## 2.1 Bases de dades:

En el camp del machine learning aplicat a la dinàmica molecular, són fonamentals les bases de dades com MD17, QM9 i altres. Aquestes bases proporcionen una gran quantitat de dades experimentals obtingudes mitjançant simulacions quàntiques utilitzant mètodes tra-dicionals, que permeten entrenar models predictius per comprendre i simular el comportament de molècules i materials a nivell atòmic.

### **2.2.1 QM9**:

Figura 1: Heatmap de la distribució de les entrades de la base de dades QM9. En l’eix d’abscisses el nombre d’elements de les molècules i per a cada fila del heatmap la distribució de les molècules amb l’element corresponent. S’aprecia que les molècules amb fluor present no segueixen la tendència general.

La base de dades QM9 és una col·lecció de dades moleculars que proporciona informació essencial sobre propietats químiques i físiques de diverses molècules orgàniques petites. Les dades de la base de dades QM9 provenen de simulacions quàntiques utilitzant mètodes tradicionals. Les simulacions s'han dut a terme per a diverses molècules orgàniques amb un nombre limitat d'àtoms. Aquestes simulacions proporcionen informació detallada sobre diverses propietats moleculars, com ara energies, capacitats calorífiques, dipols moleculars, polaritzabilitats, entre altres.

* **Detalls tècnics de les dades:**

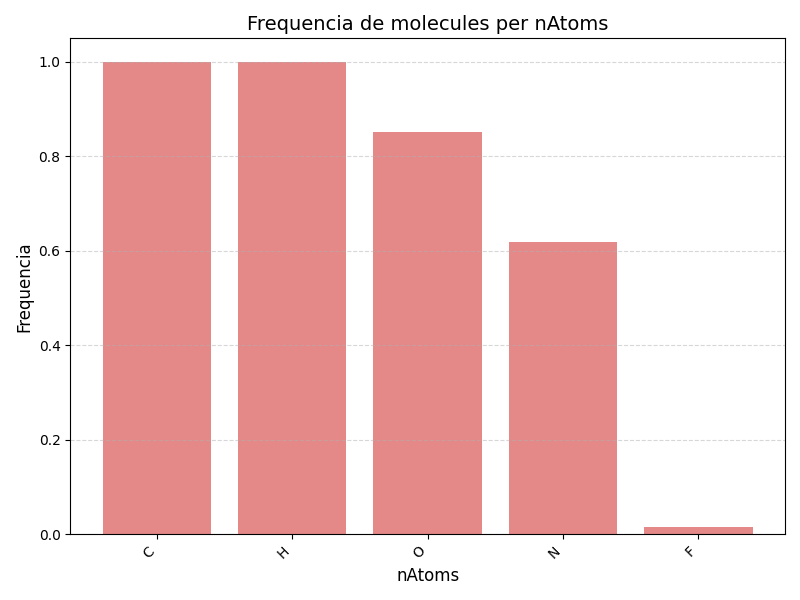
La base de dades QM9 disposa de aproximadament 134,000 molècules orgàniques, les quals tenen entre 6 i 29 àtoms cada una, que poden ser C, H, O, N i F. En les figures 2 i 3 queda representada la distribució de les molècules respecte el numero de àtoms i els àtoms que les componen. Respecte la distribució dels elements, mentre que els elements C i H apareixen a un 100% de les molècules i O i N un 80% i 60 % respectivament, el fluor apareix nomes en aproximadament 3000 molècules de les 134000 totals.

Figura 3: Histograma de la freqüència d'aparició per numero de àtoms en les molècules de la base de dades QM9.

Figura 2: Histograma del percentatge de molècules per element en la base de dades QM9.

En el cas de les dimensions de les molècules trobem una distribució normal on els números de molècules mes freqüents son 17, 18 i 19 com es pot apreciar en la Figura 1 i 3.

## 2.2 Publicacions previes:

### **2.2.1 SchNetPack:**

# 3 Conclusió

..... .... .... .......... ...... ........ ...... ..... ...... ..... ..... .... .......... ...... ........ ............ ........ ...... ..... ...... ..... ..... .... .... .......... ...... ........ ...... ..... ...... ..... ..... .... .... .......... ...... ........ ...... ..... ...... ..... ..... .... ........ ...... .

**Agraïments**

..... .... .... .......... ...... ........ ...... ..... ...... ..... ..... .... .......... ...... ........ ............ ........ ...... ..... ...... ..... ..... .... .... .......... ...... ........ ...... ..... ...... ..... ..... .... .... .......... ...... ........ ...... ..... ...... ..... ..... .... ........ ...... .

**Bibliografia**

1. Referència 1
2. Referència 2
3. Etc.

**APÈNDIX**

**A1. Secció d’apèndix**

..... .... .... .......... ...... ........ ...... ..... ...... ..... ..... .... .......... ...... ........ ............ ........ ...... ..... ...... ..... ..... .... .... .......... ...... ........ ...... ..... ...... ..... ..... .... .... .......... ...... ........ ...... ..... ...... ..... ..... .... ........ ...... .

**A2. Secció d’apèndix**

..... .... .... .......... ...... ........ ...... ..... ...... ..... ..... .... .......... ...... ........ ............ ........ ...... ..... ...... ..... ..... .... .... .......... ...... ........ ...... ..... ...... ..... ..... .... .... .......... ...... ........ ...... ..... ...... ..... ..... .... ........ ...... .