**SUMARI DE L’INFORME INICIAL**

1. **INFORMACIO PRELIMINAR / ANTECEDENTS**

La simulació de dinàmica molecular és una tècnica computacional per estudiar el comportament de sistemes moleculars a nivell atòmic o molecular. Es resolen numèricament les equacions de moviment de les partícules del sistema mitjançant mètodes d'integració numèrica, calculant les forces que actuen sobre cada partícula a partir dels potencials d'interacció entre elles. Això proporciona informació sobre l'estructura, energia, dinàmica i altres propietats del sistema.

La simulació de dinàmica molecular té un ampli ventall d'aplicacions en diferents àmbits. En la indústria farmacèutica, s'utilitza per dissenyar i optimitzar fàrmacs a partir de la comprensió de les interaccions moleculars entre els components del medicament i el receptor del pacient. En el camp de la biologia, la simulació de dinàmica molecular s'utilitza per estudiar les proteïnes i altres biomolècules, analitzant les seves propietats estructurals i la seva interacció amb altres molècules. En l'enginyeria de materials, la simulació de dinàmica molecular s'utilitza per dissenyar materials amb propietats específiques, com ara resistència mecànica o conductivitat tèrmica. A més, aquesta tècnica també s'utilitza en la física per estudiar les propietats dels sòlids i dels líquids a nivell molecular, i en la química per investigar les reaccions químiques i les propietats de les molècules.

Per resoldre les equacions de moviment en la simulació de dinàmica molecular, s'utilitzen mètodes numèrics com l'algorisme de Verlet o l'algorisme de Gear. Aquests mètodes es basen en la discretització del temps i en la resolució iterativa de les equacions de moviment per obtenir l'evolució temporal de les posicions i velocitats de les partícules. Així mateix, per calcular les forces d'interacció entre les partícules s'utilitzen diferents mètodes, com ara el mètode “*boundary element*” o el mètode de la “*particle-particle*”. L'elecció del mètode numèric depèn de la naturalesa del sistema i de la precisió i eficiència que es requereixen en la simulació.

A mesura que augmenta la complexitat del sistema, augmenta també el nombre de molècules i la quantitat de dades que cal processar. Això, a la seva vegada, augmenta el temps necessari per a completar cada pas de temps. Així, per exemple, per a simular la dinàmica molecular d'una proteïna amb milions d'àtoms, són necessàries simulacions que poden requerir desenes de milers de passos de temps i que poden trigar dies o setmanes per a ser completades en un ordinador de propòsit general. Per això, els investigadors de la dinàmica molecular han hagut de recórrer a tècniques d'optimització i paral·lelització per a reduir el temps necessari per a completar les simulacions. Això inclou l'ús de supercomputadors i la programació en paral·lel, que permeten repartir el treball en diferents nodes d'un sistema computacional per a accelerar el procés.

En resum, els mètodes numèrics utilitzats en la dinàmica molecular són computacionalment intensius i poden requerir molt de temps per a completar simulacions de sistemes biològics complexos. La reducció del temps necessari per a aquestes simulacions continua sent un repte important per a la investigació en aquesta àrea.

Recentment, s'han desenvolupat noves tècniques basades en l'aprenentatge computacional per a fer front als problemes de complexitat de la dinàmica molecular. Aquestes tècniques inclouen els mètodes basats en xarxes neuronals, que permeten aprendre models a partir de dades experimentals o de simulacions prèvies. També s'han desenvolupat algoritmes d'aprenentatge reforçat, que poden optimitzar la selecció de paràmetres en la simulació de dinàmica molecular.

El projecte proposa fer la revisió dels treballs publicats en camps similars i fer-ne una adaptació/millora de les metodologies aplicades en aquests per poder-les aplicar en al camp de la simulació de les interaccions moleculars (a partir d’ara MD).

1. **OBJECTIU DEL TFG**

Estudiar i proposar possibles millores als mètodes utilitzats actualment per aplicar les NN a problemes de MD.  
Definir, entrenar i validar una arquitectura de NN capaç de dur a terme simulacions de MD entre molècules senzilles.

1. **METODOLOGIA**

Separaré el projecte en etapes per tal de poder estructurar millor el procés i la carga de treball, es dividirà en 4 fases, tres obligatòries i la quarta opcional (si hi ha temps a realitzar-la).

* + **Etapa 1 (Investigació preliminar):** primer de tot haurem de determinar quin/s dels treballs utilitzaren com a punt de partida, partim de 3 treballs que he escollit donat a la proximitat a l’idea inicial que tenia per el TFG.
  + **Etapa 2 (Estudi profund**): una vegada escollits els treballs, fer una lectura i anàlisis mes profund i una prova del codi per tal de entendre el funcionament d’aquests i realitzar un informe per extreure els punts principals.
  + **Etapa 3 (Millores/canvis):** a partir de l’informe realitzat, fer recerca externa als treballs en qüestió per a poder millorar (si es possible) tant el tractament previ de les dades com la estructura de les NN aplicades.
  + **Etapa 4 (Innovació):** ultima fase (opcional) buscar aplicar les idees pròpies utilitzant com a base els models estudiats, com ara la desratització de espai a l’hora de representar les simulacions.

1. **PLANIFICACIO**

**Setmana 1-2:**

* + Selecció dels tres treballs de referència i lectura preliminar.
  + Lectura i anàlisi més profund dels tres treballs de referència, realitzant un informe per extreure els punts principals.

**Setmana 3-4:**

* + Continuar amb l'anàlisi dels tres treballs de referència i prova del codi per tal de comprendre el funcionament d'aquests.
  + Iniciar la recerca externa als treballs de referència per a millorar el tractament de les dades i la estructura de les NN aplicades.

**Setmana 5-8:**

* + Continuar amb la recerca externa i realitzar les millores/canvis necessaris.
  + Analitzar amb més profunditat les diferents tècniques de xarxes neuronals utilitzades en els treballs de referència.

**Setmana 9-12:**

* + Experimentar amb diferents tècniques i paràmetres per a la millora dels resultats.
  + Començar a explorar idees per a la innovació.

**Setmana 13-16:**

* + Continuar explorant idees per a la innovació i decidir si són factibles.
  + Aprofundir en les diferents arquitectures de les xarxes neuronals.

**Setmana 17-20:**

* + Provar amb noves tècniques de tractament de dades i millores en les NN aplicades.
  + Preparació de la presentació del treball i finalització de la documentació.

1. **ANALISIS “State of the art”:**

**SchNetPack:**

Presenta una arquitectura per representar i entrenar un model per predicció de MD, La primera etapa consta de dos possibles una de cada un dels tipus mencionats anteriorment , wACSF (description based) i SchNet (end-to-end).

wACSF es una representació que es calcula prèvia al entrenament/predicció del model i esta basada en per àtom comprova per els àtoms adjacents les posicions (radial symetry functions) i els angles entre ells (angular symetry functions) i convina aquesta informació així generant uns valors per a poder representar la seva situació en l’espai.

SchNet es una DNN que aprèn a representar els sistemes a partir de un entrenament, aquesta busca seguir correctament les lleis físiques.

A l’hora de la predicció usa unes capes denses definides en el treball com a atomisme (també usades durant la representació SchNet).

Diagrama

Descripción generada automáticamente

Com a afegit també hi ha una breu explicació de com usar els models i entrenar-los a partir de bases de dades publiques amb l’ajuda de funcions (les quals no entra massa en com funcionen)

**TorchMD:**

TorchMD és un marc de treball de deep learning per a simulacions de MD. Està construït sobre PyTorch i ofereix una sèrie d'eines i models per simular eficientment el comportament de les molècules en entorns complexos.

Aquest parteix de la representació implementada en SchNet però re-implementa per complet tant la part de entrenament com la de inferència per permetre utilitzar les eines que proporciona el marc de treball de PyTorch com ara l’entrenament en múltiples GPUs paral·lelament. A l’hora d’entrenar utilitza el PyTorch lightning framework [[Link](https://github.com/Lightning-AI/lightning)].

Per acabar aquest treball destaca en les aplicacions i resultats que presenta, du a terme 4 experiments:

* + Comparació dels resultats amb ACEMD3
  + Entrenament i avaluació en la base de dades QM9
  + Demostració de la utilitat del model end-to-end (SchNet)
  + Coarse-Graning (baix nivell de detall) en sistemes de “proteine folding”

**DeePMD-kit:**

Es un paquet de programari de codi obert que utilitza DNN per a la construcció de potencials d'interacció atòmica es a dir és capaç de predir les forces i les energies que actuen sobre els àtoms d'un sistema en funció de la seva configuració espacial. Aquest treball esta enfocat a la simulació de TiN el qual una estructura cristal·lina cúbica (que per tant no s’adhereix gaire a l’objectiu del treball).

Diagrama

Descripción generada automáticamenteSegueix un funcionament similar a SchNet en el que es defineixen uns descriptors, a partir d’una funció que te en compte tant les coordenades com altres factors s’obtenen uns valors els quals passen per una capa de embedding NN que s’ocupa d’acabar de filtrar-los i per acabar es tornen a combinar amb les coordenades per constituir els descriptors que s’utilitzaran per la predicció.

El treball també entra en detall de el software usat per a generar les dades d’entrenament el qual es CALPHAD, Thermo-Calc 2021b i TCTI3 (en aquest no explica el funcionament intern d’aquests softwares però pot ser útil poder consultar-los en cas de necessitar generar dades)

Per acabar fa una sèrie de tests on compara el model NNP amb el software MEAM (el mètode establert actualment per fer aquest tipus de simulacions) i observa que NNP obté resultats molt similars a MEAM amb un cost i una eficiència molt menor per tant fent el mètode un èxit en aquest camp.

1. **Bibliografia:**

[1] - K. T. Schütt, P. Kessel, M. Gastegger, K. A. Nicoli, A. Tkatchenko, and K.-R. Müller. **SchNetPack:** **A Deep Learning Toolbox For Atomistic Systems**. *Journal of chemical theory and computation*, 2019, 15, 448-455. [Link](https://pubs.acs.org/doi/10.1021/acs.jctc.8b00908)

[2] - Stefan Doerr, Maciej Majewski, AdriàPérez, Andreas Krämer, Cecilia Clementi, Frank Noe, Toni Giorgino, and Gianni De Fabritiis. **TorchMD: A Deep Learning Framework for Molecular Simulations**. *Journal of chemical theory and computation*, 2021, 17, 2355−2363. [Link](https://pubs.acs.org/doi/10.1021/acs.jctc.0c01343)

[4] - Takeru Miyagawa, Kazuki Mori, Nobuhiko Kato , Akio Yonezu. **Development of neural network potential for MD simulation and its application to TiN.** *Computational Material Sciencie*, 15 April 2022, 111303. [Link](https://www.sciencedirect.com/science/article/abs/pii/S0927025622000970)