

# INFORME INICIAL

## 1 INFORMACIO PRELIMINAR / ANTECEDENTS

---

La simulació de dinàmica molecular (a partir d'ara DM) és una tècnica computacional que estudia el comportament de sistemes moleculars, proporcionant informació sobre l'estructura, l'energia, la dinàmica i altres propietats del sistema.

La simulació de DM 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 (el pacient). En el camp de la biologia, la simulació de DM 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 DM 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 DM, s'utilitzen mètodes numèrics com l'algorisme de Verlet<sup>[1][2]</sup> o l'algorisme de Gear<sup>[3]</sup>. 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"<sup>[4]</sup>. 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 DM d'una proteïna amb milions d'àtoms, són necessàries simulacions que poden requerir desenes de milers d'iteracions i que poden trigar dies o setmanes per a ser completades en un ordinador convencional. Per això, els investigadors de la DM 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 DM són computacionalment intensius i poden requerir molt de temps per a completar simulacions de sistemes 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 DM<sup>[5][6][7]</sup>. Aquestes tècniques inclouen els mètodes basats en xarxes neuronals (a partir d'ara XN), que permeten entrenar models a partir de dades experimentals o de simulacions prèvies. Aquest Treball proposa estudiar els estudis publicats i fer una adaptació/millora de les metodologies aplicades en aquests.

## 2 OBJECTIU DEL TFG

En aquesta secció, s'estableixen els objectius del projecte i se'ls assigna una prioritat segons la seva importància i grau d'impacte en el resultat final. Per a això, es proposa la utilització d'una taula que defineixi clarament cada objectiu i la seva categoria de prioritat, que pot ser essencial, secundari o opcional. Així, es podrà establir una guia clara per al desenvolupament del projecte i es podrà assegurar que s'assoleixen els objectius més importants i necessaris en primer lloc.

Taula 1 – Sumari dels objectius del treball

Objectiu	Descripció	Prioritat
O1	Desenvolupar un coneixement profund dels fonaments teòrics i pràctics de les <b>XN</b> i la <b>simulació de DM</b>	Essencial
O2	Desenvolupar una comprensió crítica dels avantatges i les limitacions de les XN en la simulació de DM, comparant-les amb altres <b>tècniques i abordatges existents</b> .	Essencial
O3	<b>Estudiar</b> els treballs mes recents del camp de la DM que apliquen XN.	Essencial
	O3.1 Bases de dades utilitzades	
	O3.2 Mètodes de representació dels sistemes de molècules	
	O3.3 Arquitectures de XN utilitzades	
	O3.4 Avaluar el rendiments dels mètodes	
O4	<b>Explor noves estratègies</b> per millorar els mètodes usats en els treballs estudiats.	Essencial
	O4.1 Proposar noves fonts de dades d'entrenament	
	O4.2 Dissenyar noves arquitectures de XN	
	O4.3 Explorar tècniques de transferència d'aprenentatge	
	O4.4 Comparar el rendiments amb els mètodes originals	
O5	<b>Definir, entrenar i validar</b> una arquitectura de XN específicament dissenyada per a la simulació de DM per interaccions <b>entre molècules senzilles</b> .	Secundari
	O4.1 Escollir bases de dades a utilitzar per l'entrenament i validació	
	O4.2 Definir mètodes de representació dels sistemes de molècules	
	O4.3 Explorar diferents tipus d'arquitectures de XN	
	O4.4 Comparar el rendiments amb els dels altres treballs	

### 3 METODOLOGIA

Prenent de referència i simplificant la metodologia àgil utilitzada regularment en el camp del desenvolupament software, se separarà el projecte en fases. En cada fase es poden establir cicles de treball de durada flexible (entre una i dos setmanes) per assegurar un seguiment adequat del progrés del projecte. Per a cada cicle, es predefiniran objectius específics (Taula 1) a assolir i es generaran informes de cicle per verificar si s'han assolit tots els objectius desitjats i explicar les raons si no s'han assolit. També es permetrà canviar l'ordre dels cicles de treball sempre que es justifiqui adequadament o no afecti negativament a altres tasques pendents. Al finalitzar cada fase es redactarà un informe de progrés que recollirà els continguts dels informes dels cicles que componen la fase.

Taula 2 – Fases del Treball

<b>Fase de formació:</b> En aquesta fase s'ha de desenvolupar un coneixement profund dels fonaments teòrics i pràctics de les XN i la simulació de DM. Aquesta fase es pot realitzar amb una combinació de lectures i tutories amb especialistes dels temes.
<b>Fase d'avaluació:</b> En aquesta fase es realitzarà una avaluació crítica dels avantatges i les limitacions de les XN en la simulació de DM, comparant-les amb altres tècniques i abordatges existents. Es realitzaran estudis de casos específics en els quals s'utilitzin XN per a la simulació de DM, i es valorarà el seu rendiment.
<b>Fase d'exploració:</b> En aquesta fase es proposaran noves estratègies per millorar els mètodes utilitzats en els treballs estudiats. Es poden proposar noves fonts de dades d'entrenament, dissenyar noves arquitectures de XN i explorar tècniques de transferència d'aprenentatge. S'ha de comparar el rendiment de les noves estratègies amb els mètodes originals.
<b>Fase d'innovació:</b> En aquesta fase es definirà, entrenarà i validarà una arquitectura de XN específicament dissenyada per a la simulació de DM per interaccions entre molècules senzilles. Es poden escollir bases de dades per a l'entrenament i la validació, definir mètodes de representació dels sistemes de molècules i explorar diferents tipus d'arquitectures de XN. Es compararà el rendiment de la nova arquitectura amb altres treballs estudiats.

Aquestes fases seran realitzades al llarg de 20 setmanes, tal com s'explica en el següent apartat del document (Taula 3).

Es duran a terme reunions amb el tutor del projecte en efectuar l'entrega de cada informe de progrés, també es podran fer reunions suplementàries tant amb el tutor "intern" com amb el tutor "extern" segons el que es consideri més apropiat.

## 4 PLANIFICACIÓ

La taula de planificació (Taula 3) inclou una sèrie d'objectius específics que han de ser assolits en un període de temps determinat. A través d'aquests objectius, es pretén comprendre les limitacions i avantatges de les XM, recopilar informació dels treballs existents, identificar les tècniques de XN més efectives i aplicar-les per crear noves arquitectures de XN específiques. A més a més, la planificació també inclou l'elaboració de dos informes de progrés, un informe final i la preparació d'una presentació i un pòster per comunicar els resultats obtinguts.

Taula 3: Planificació setmanal

Setmana	Objectius	Assoliment
1	O1	Comprensió general dels conceptes clau
2	O1, O2	Identificació de les limitacions de les XN en DM
3	O2	Comprensió dels avantatges de les XN en DM
4	O2, O3	Recopilació d'informació dels treballs existents
<b>Informe Inicial</b>		
5	O3.1, O3.2	Conèixer les bases de dades disponibles
6	O3.2, O3.3, O3.4	Entendre les arquitectures de les XN aplicades
7	O3.2, O3.3, O3.4	Estimar el rendiment de les XN
8	O3.3, O3.4	Comprendre les tècniques XN més efectives
9	O3.4, O4.1	Identificar les tècniques XN més eficients
<b>Informe de Progrés (I)</b>		
10	O4.1	Explorar noves fonts de dades
11	O4.1, O4.2	Creació de noves arquitectures de XN
12	O4.1, O4.2 O4.3	Creació de noves arquitectures de XN
13	O4.2 O4.3, O4.4	Aplicar tècniques de transferència d'aprenentatge
14	O4.3 O4.4	Comparar el rendiment amb els originals
<b>Informe de Progrés (I)</b>		
15	O5.1, O5.2	Seleccionar dades per a l'entrenament
16	O5.1, O5.2, O5.3	Creació de noves arquitectures XN específiques
17	O5.2, O5.1, O5.4	Comparar rendiments amb altres treballs
<b>Informe final</b>		
18	-	Preparació Presentació
19	-	Preparació Presentació
<b>Proposta de Presentació</b>		
20	-	Realització del Pòster
<b>Pòster</b>		

## 5 BIBLIOGRAFIA

---

- [1] - Frenkel, Daan, and Berend Smit. **Understanding Molecular Simulation from Algorithms to Applications**. *Academic Press*, 2002.
- [2] - Braun, Efrem, et al. **Best Practices for Foundations in Molecular Simulations [Article v1.0]**. *Living Journal of Computational Molecular Science*, vol. 1, no. 1, 2019
- [3] - M. P. Allen and D. J. Tildesley. **Computer Simulations of Liquids**. *Oxford University Press*, 2002
- [4] - Giordano Mancini, Marco Fusè, Filippo Lipparini, et al. **Molecular Dynamics Simulations Enforcing Nonperiodic Boundary Conditions: New Developments and Application to the Solvent Shifts of Nitroxide Magnetic Parameters**. *Journal of Chemical Theory and Computation*. 2022, 18, 4, 2479–2493.
- [5] - K. T. Schütt, P. Kessel, M. Gastegger, et al. **SchNetPack: A Deep Learning Toolbox For Atomistic Systems**. *Journal of chemical theory and computation*, 2019, 15, 448-455.
- [6] - Stefan Doerr, Maciej Majewski, Adrià Pérez, et al. **TorchMD: A Deep Learning Framework for Molecular Simulations**. *Journal of chemical theory and computation*, 2021, 17, 2355–2363.
- [7] - Takeru Miyagawa, Kazuki Mori, Nobuhiko Kato , et al. **Development of neural network potential for MD simulation and its application to TiN**. *Computational Material Science*, 2022, 15 April, 111303.