

Potentiels répulsifs transférables pour la méthode Density Functional based Tight-Binding: approche novatrice par apprentissage automatique

Dylan Bissuel

► To cite this version:

Dylan Bissuel. Potentiels répulsifs transférables pour la méthode Density Functional based Tight-Binding: approche novatrice par apprentissage automatique. Physique [physics]. Université de Lyon, 2021. Français. NNT: 2021LYSE1276. tel-03704165

HAL Id: tel-03704165 https://theses.hal.science/tel-03704165

Submitted on 24 Jun2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.





THÈSE de DOCTORAT DE L'UNIVERSITÉ DE LYON Opérée au sein de : l'Université Claude Bernard Lyon 1

Ecole Doctorale 52 Physique et Astrophysique

Spécialité de doctorat : Physique

Soutenue publiquement le 17/12/2021, par : $\mathbf{Dylan \ Bissuel}$

Potentiels répulsifs transférables pour la méthode Density Functional based Tight-Binding, approche novatrice par apprentissage automatique

Devant le jury composé de :	
Abdul-Rahman Allouche	Président
Professeur, Université Claude Bernard Lyon 1	
Thierry Deutsch	Rapporteur
Directeur de recherche, MEM, CEA Grenoble	
Mathias Rapacioli	Rapporteur
Chargé de recherche, LCPQ, UMR5626, Toulouse	
Magali Benoit	Examinatrice
Chargée de recherche, CEMES, UPR8011, CNRS Toulouse	
Sandrine Brochard	Examinatrice
Professeure, Université de Poitiers	
Aude Simon	Examinatrice
Chargée de recherche, LCPQ, UMR5626, Toulouse	
Thomas Niehaus	Directeur de thèse
Professeur, Université Claude Bernard Lyon 1	
Tristan Albaret	Co-directeur de thèse
Maître de conférences, Université Claude Bernard Lyon 1	
Emeric Bourasseau	Invité
Chef de laboratoire, CEA Cadarache	

Potentiels répulsifs transférables pour la méthode Density Functional based Tight-Binding, approche novatrice par apprentissage automatique

Dylan BISSUEL







Modélisation de la matière condensée et des interfaces, institut Lumière Matière ${\rm UMR5306-UCBL-CNRS}$

Résumé

Cette thèse propose une amélioration pour la méthode *Density Functional based Tight-Binding* (DFTB) en se concentrant sur les potentiels répulsifs, que l'on représente comme des interactions à plusieurs corps à l'aide de réseaux de neurones de haute dimension couplés à des fonctions de symétrie de Behler-Parrinello.

La première partie de ces travaux en présente le cadre théorique et les outils conceptuels principaux. À partir d'un aperçu succinct des méthodes premier principe, et en particulier de la théorie de la fonctionnelle de la densité (DFT), le formalisme *DFTB* est détaillé avec une emphase particulière sur le terme de l'énergie répulsive. Par la suite, une présentation générale des outils d'apprentissage automatique est donnée, avant de s'intéresser plus spécifiquement aux applications en simulation atomistique et aux *potentiels numériques* en résultant.

La deuxième partie s'inscrit dans la continuité directe de la précédente en proposant une méthodologie permettant de greffer et d'utiliser ces mêmes outils au sein du formalisme DFTB. Plusieurs grands axes en sont présentés : l'obtention de données DFTfaisant office de références (avec application au Silicium pur), ainsi que l'implémentation et l'entraînement de réseaux de neurones pour les reproduire.

La dernière partie rend compte quantitativement des apports *in-situ* des potentiels répulsifs numériques dans le code DFTB+. Celle-ci se focalise sur la qualité de représentation pour une variété de systèmes d'intérêt du Silicium pur, en identifiant des caractéristiques clés et en les comparant conjointement à la méthode DFT de référence et aux potentiels répulsifs de paire utilisés par DFTB jusqu'alors. La capacité à représenter à la fois les systèmes périodiques et isolés est notamment étudiée. Nous discutons également de l'intérêt, de la performance et des limitations de notre approche par rapport à deux potentiels complètement paramétriques.

On obtient au final un potentiel répulsif fiable, reproduisant fidèlement les propriétés de référence des phases cristallines et désordonnées, ce qui constitue une amélioration sensible par rapport aux potentiels de paire. À cause de difficultés sporadiques à converger quelques géométries, traduites par des forces interatomiques parfois surestimées, la performance globale sur les systèmes isolés est globalement inférieure aux potentiels de paire; ces derniers, plus fiables, restent préférables pour les nano-objets. La comparaison aux potentiels purement numériques étudiés est très disputée : ceux-ci font preuve d'une flexibilité plus marquée, synonyme bonnes performances interpolatoires mais aussi d'une robustesse physique moindre.

Mots-clés

Simulation atomistique, apprentissage automatique, Density Functional based Tightbinding, potentiels répulsifs, réseaux de neurones de haute dimension, fonctions de symétrie centrées sur les atomes de Behler-Parrinello, Silicium, monocristaux, liquides, amorphes, nano-clusters.

Transferable repulsive potentials for DFTB : a novel machine-learning approach

Abstract

This thesis focuses on improving the *Density Functional based tight-binding* method through the so-called repulsive potential. We represent it as a many-body interaction with high-dimensional neural-networks and *Behler-Parrinello* symmetry functions.

First, we define the theoretical framework and conceptual tools. Starting from a quick overview of first principle methods, and especially *Density Functional Theory* (DFT), the *DFTB* formalism is presented with emphasis on the repulsive energy. Then, a global presentation of machine-learning is provided, before narrowing our interest to specific applications coupling it with atomistic simulation, hence focusing on numerical potentials they generate.

The next part is in the direct continuity of the former, proposing a methodology allowing to use those very tools within the DFTB formalism. Several directions are explored : DFT calculations to obtain a set of meaningful reference data (applied to pure silicon) and the implementation and training of neural-networks to reproduce such data.

Eventually, we evaluate *in-situ* the performance of such numerical repulsive potentials with the DFTB+ software. We focus on representing important systems for pure silicon; we identify key features and compare it to both DFT reference and the pair repulsive potential used until now in DFTB. In particular, we study the ability to cover periodic and non-periodic systems with the same tool. We also discuss the advantages, perfomance and limitations of our approach compared to two purely numerical potentials.

We get a reliable potential, reproducing accurately the properties of crystals and disordered systems, which is a significant improvement compared to pair repulsive potentials. Because of struggles to sometimes optimize geometries of isolated objects, seen through an overestimation of interatomic forces, the overall performance on such systems is inferior to pair potentials; those are indeed more trustworthy, and should therefore be preferred for nano-clusters. The comparison to fully-numerical potentials provides mixed results : those show a greater flexibility than our tool, which entails both good interpolation performance and lower physical robustness.

Keywords

Atomistic simulation, machine-learning, Density Functional based Tight-binding, repulsive potentials, high-dimensional neural-networks, Behler-Parrinello atom-centered symmetry functions, Silicium, monocristals, liquids, amorphous, nano-clusters.

Table des matières

1	Intr	roduction	11
2	Potentiels premier principe		
	2.1	Introduction historique	14
	2.2	Théorie de Hartree-Fock	15
	2.3	Théorie de la fonctionnelle de la densité	20
		2.3.1 Approximation de la densité locale	22
		2.3.2 Approximation du gradient généralisé	23
3	La	méthode Density-Functional based Tight-Binding	25
	3.1	Développement au premier ordre	27
	3.2	Développement au second ordre	29
	3.3	Énergie répulsive	30
4	Obj	ectif	36
5	App	prentissage automatique	42
	5.1	Présentation historique	42
	5.2	Applications en simulation atomistique	43
	5.3	Outils d'apprentissage	46
		5.3.1 Réseaux de neurones	46
		5.3.2 Méthodes à noyaux	53
	5.4	Décrire l'environnement atomique	55
		5.4.1 Fonctions de symétrie centrées sur les atomes	57

		5.4.2	Autres descripteurs	<i>i</i> 0
			5.4.2.1 Bispectre de la densité des voisins 6	51
			5.4.2.2 Matrices de Coulomb	53
6	App	olicatio	n aux potentiels répulsifs 6	5
	6.1	Calcul	de l'énergie répulsive	i5
	6.2	Optim	ser un réseau de neurones 6	57
		6.2.1	Rétropropagation du gradient	58
			6.2.1.1 Apprentissage sur l'énergie répulsive	'0
			6.2.1.2 Apprentissage sur les forces répulsives	'2
		6.2.2	Filtre de Kalman étendu	'6
7	Cor	nstruct	on d'un potentiel répulsif numérique 7	'9
	7.1	Généra	tion des données de référence	'9
		7.1.1	Choix des systèmes	' 9
		7.1.2	Calculs de structure électronique	31
			7.1.2.1 Calculs VASP	32
			7.1.2.2 Calculs DFTB+ \ldots 88	34
		7.1.3	Translation des énergies répulsives	35
		7.1.4	Sélection des données de référence	37
		7.1.5	Le cas des très faibles distances	39
	7.2	Appre	tissages)1
		7.2.1	Apprentissage sur des systèmes simples)3
		7.2.2	Ajout de configurations liquides)5
		7.2.3	Ajout des courbes énergies vs. volume)6
	7.3	Résun	é et procédure à suivre 10)()
8	App	oort de	s potentiels répulsifs numériques 10	3
	8.1	Crista	x de Silicium pur)5
		8.1.1	Stabilité relative)5
		8.1.2	Courbes énergie vs. volume)7

		8.1.3	Densités d'états vibrationnels	110
		8.1.4	Défauts ponctuels du Silicium diamant	113
	8.2	Nano-	clusters de Silicium pur	117
		8.2.1	Accord global avec VASP dans la prédiction des configurations stable	es117
		8.2.2	Prédiction de la compacité des nano-clusters	121
	8.3	Systèn	nes désordonnés de Silicium pur	125
		8.3.1	Silicium liquide	126
		8.3.2	Configurations amorphes	130
9	Con	clusio	n	135
\mathbf{A}	nnex	eAC	hoix des descripteurs	149
	A.1	Généra	ation de jeux de paramètres	149
	A.2	Sélecti	on préalable	152
	A.3	Sélecti	on après apprentissage	153
\mathbf{A}	nnex	e B C	alcul des forces répulsives	155
	B.1	Dérivé	es d'un réseau de neurones	156
	B.2	Dérivé	es des fonctions de symétrie	159
		B.2.1	Dérivées des fonctions de symétrie radiales	160
		B.2.2	Dérivées des fonctions de symétrie angulaires	161
\mathbf{A}	nnex	e C In	clusion des outils d'apprentissage automatique dans $DFTB$ -	+166
	C.1	Intégra	ation dans $DFTB+$	166
		C.1.1	Fonctionnement de $DFTB+$	166
		C.1.2	Vue générale de la librairie et interfaçage	167
			C.1.2.1 Structure du code	168
			C.1.2.2 Normalisation	169
			C.1.2.3 Interfaçage à $DFTB+$	172
	C.2	Utilisa	tion et performance	174
\mathbf{A}	nnex	e D D	iagramme UML de la librairie	179

-
E
<u> </u>
\sim

Annexe E Résultats supplémentaires : approches purement numériques 181

Remerciements

185

Table des figures

3.1	Vue générale des méthodes pour représenter la matière et des échelles as- sociées	26
3.2	Calcul d'un potentiel répulsif de paire pour la liaison C-C	32
3.3	Évaluation du potentiel répulsif de paire pour le Silicium utilisé dans cette thèse	33
4.1	Exemple de détermination de l'énergie répulsive d'une liaison C-H dans la molécule de méthane	36
4.2	Structure cubique face centrée et distance aux premiers et seconds voisins pour l'or	37
5.1	Présentation schématique du concept de potentiel numérique	45
5.2	Schéma du fonctionnement d'un neurone artificiel	47
5.3	Différence entre les deux classes de réseaux de neurones	47
5.4	Exemple d'utilisation courante de réseaux de neurones récurrents	48
5.5	Schéma de fonctionnement d'un réseau de neurones de forme 4-2-3-1	49
5.6	Schématisation du phénomène de sur-apprentissage	51
5.7	Détecter le sur-apprentissage.	52
5.8	Extraction de caractéristiques - exemple de MNIST	55
5.9	Représentation schématique d'un environnement atomique	56
5.10	Exemples de fonctions de symétrie G_2	59
5.11	Partie angulaire de quelques fonctions de symétrie G_4	60
5.12	Projection de Riemann	62
5.13	Concept d'environnement de liaison atomique	63
6.1	Calcul de l'énergie répulsive d'un environnement atomique	66

6.2	Calcul de l'énergie répulsive d'une boîte de simulation. \ldots . \ldots . \ldots	67
6.3	Schématisation de l'impact du taux d'apprentissage	70
7.1	Schématisation du processus de translation des énergies répulsives de ré- férence.	86
7.2	Exemple de fonction de distribution radiale, construite comme un histo- gramme	88
7.3	Schéma du fonctionnement du potentiel répulsif <i>ad-hoc</i> de courte portée.	91
7.4	Processus d'apprentissage : réseau global avec un jeu de données de référence d'une complexité modérée	94
7.5	Processus d'apprentissage : réseau global avec un jeu de données de référence d'une complexité intermédiaire	95
7.6	Prédiction des énergies de liaison en fonction du volume pour les cristaux de Silicium pur par un réseau inadéquat.	97
7.7	Processus d'apprentissage : réseau global avec un jeu de données de référence d'une complexité élevée	99
8.1	Écarts en énergie des différents cristaux au Silicium diamant	106
8.2	Stabilité relative des différents cristaux de Silicium pur en fonction du volume pour <i>DFTB</i> avec différents potentiels répulsifs	108
8.3	Énergie répulsive des monocristaux en fonction de la distance au plus proche voisin.	109
8.4	Densités d'états vibrationnels des cristaux de Silicium pur pour différents potentiels répulsifs <i>DFTB</i>	111
8.5	Défauts interstitiels du Silicium diamant	114
8.6	Comparaison des énergies de formation de défauts ponctuels	115
8.7	Comparaison des accords avec DFT dans la prédiction des minima d'énergie de configurations de nano-clusters optimisées par $VASP$	118
8.8	Comparaison des accords avec <i>DFT</i> dans l'optimisation et la prédiction des minima d'énergie	119
8.9	Évolution des énergies et des forces pour l'optimisation des géométries de deux nano-clusters Si_{14} .	120
8.10	Rayons de giration des nano-clusters de Silicium.	122
8.11	Énergies répulsives par atome calculées pour plusieurs nano-clusters Si_{30} .	124

8.12	Fonctions de distribution radiale du Silicium liquide à différentes tempé- ratures	127
8.13	Fonctions de distribution angulaire du Silicium liquide à différentes tem- pératures	129
8.14	Densité d'états vibrationnels de plusieurs systèmes amorphes de Silicium pour différents potentiels répulsifs <i>DFTB</i>	133
A.1	Fonctions de symétrie radiales générées automatiquement	151
C.1	Sélection d'images périodiques pour reconstruire la liste de voisins	174
C.2	Comparaison des temps de calcul $DFTB+$ de différents réseaux de neurones par rapport aux interactions de paires	176
C.3	Comparaison des temps de calcul de différents réseaux de neurones par rapport aux interactions de paires : systèmes plus modestes	177
E.1	Stabilité relative des différents cristaux de Silicium pur en fonction du volume avec des potentiels purement paramétriques	182
E.2	Densités d'états vibrationnels des cristaux de Silicium pur pour des poten- tiels purement numériques.	183
E.3	Densité d'états vibrationnels de plusieurs systèmes amorphes de Silicium pour des potentiels purement numériques.	184

Liste des tableaux

7.1	Erreurs sur les énergies et les forces : réseau global avec un jeu de données de référence d'une complexité modérée	94
7.2	Erreurs sur les énergies et les forces : réseau global avec un jeu de données de référence d'une complexité intermédiaire.	96
7.3	Erreurs sur les énergies et les forces : réseau global avec un jeu de données de référence d'une complexité élevée.	98
8.1	Caractéristiques structurelles des amorphes relaxés par chaque méthode.	131

Chapitre 1

Introduction

Dès le paléolithique, la volonté de nos ancêtres de connaître et comprendre le monde alentour émerge avec l'apparition de formes sommaires de mathématiques et d'astronomie : la science trouve alors sa place aux côtés de la technique, et n'aura dès lors cesse de l'alimenter. Au-delà des nombreux changements de paradigmes connus par les civilisations successives, les scientifiques de tous temps ont rivalisé d'ingéniosité pour développer des modèles théoriques rendant compte des observations expérimentales, et pour éprouver à l'inverse des prédictions théoriques à coup de dispositifs expérimentaux pertinents. Ces deux pendants ont d'ailleurs tout à gagner à ne pas s'exclure l'un l'autre. Par exemple, Ptolémée avance à tort dans son Almageste au II^e siècle qu'un objet lâché depuis le haut du mât d'un navire en mouvement ne chutera pas vers la base du mât, mais vers l'arrière du vaisseau. Cette assertion aspire à soutenir le modèle, alors dominant, d'une Terre immobile et d'un Univers géocentrique. Figure d'autorité justifiée par son statut de dépositaire du savoir accumulé au cours de l'Antiquité, son modèle alimente pour plus d'un millénaire la croyance d'une Terre au centre de tout; faire l'expérience lui aurait permis - ou à ses nombreux successeurs - d'en constater l'invalidité, et aurait sans doute accéléré nos progrès occidentaux en astronomie.

Des centaines d'années plus tard, les sciences s'emparent de l'outil informatique très tôt après son avènement au cours du XX^e siècle. Elles y voient une opportunité certaine pour combler les incomparabilités entre théorie et expérience : modèles théoriques trop lourds pour étudier des systèmes réels à la main, échelles de taille ou de temps inaccessibles aux expérimentateurs, etc. À la dualité déca-millénaire entre théorie et expérience vient donc s'ajouter la simulation numérique. Le faible coût relatif de l'utilisation d'outils numériques en comparaison du matériel expérimental en fait une alternative intéressante. Loin de s'opposer aux manipulations expérimentales, ils permettent de faire la jonction entre théorie et expérience, en se confirmant ou s'infirmant mutuellement afin d'enrichir les modèles existants et de pouvoir corriger, consolider ou étendre notre connaissance du monde qui nous entoure.

L'étude des matériaux en constitue un exemple très convaincant : à l'échelle macroscopique, nous disposons d'un panel extensif de méthodes expérimentales fiables permettant de caractériser un matériau, et les modèles théoriques empiriques aspirent donc à reproduire l'expérience. A l'inverse, les très petites échelles sont régies par la mécanique quantique : le modèle théorique est connu, mais bien trop complexe pour s'intéresser à des systèmes réels et des approximations sont nécessaires. Dans le même temps, les expériences sont plus ardues à contrôler, conduire et interpréter. Ici, l'outil numérique permet d'une part de créer des expériences *in-silico* parfaitement contrôlées, et d'autre part d'approximer plus ou moins finement les équations de la mécanique quantique pour rendre des systèmes de plus en plus pertinents accessibles, facilitant par la même occasion la comparaison avec les résultats expérimentaux.

Lors de la dernière décennie, le paysage de la simulation des matériaux a été fortement modifié par l'utilisation de nombreux outils d'apprentissage automatique. Bien que la majorité des chercheurs sondent évidemment des applications variées, tous orientent leur travaux dans une même direction : redéfinir le compromis entre précision et coût numérique. Dans l'idée, ceci n'a rien de surprenant : la volonté d'étendre la précision des méthodes premier principe à des systèmes toujours plus grands est au cœur de la discipline depuis ses balbutiements. L'apprentissage automatique est dès lors un formidable tremplin, et représente un bond d'une ampleur inédite offrant des perspectives prometteuses. En revanche, dans le cas des outils purement numériques le gain de performance va souvent de paire avec une perte d'information (par exemple l'accès à la structure électronique) ou avec l'apparition d'une dépendance forte à des méthodes purement mathématiques, dont les fondations physiques sont très réduites.

Les méthodes semi-empiriques sont bien plus anciennes, comme la méthode *Density Functional based Tight-Binding* (DFTB) apparue à la fin du siècle dernier. Elles partagent pourtant la même volonté que les outils d'apprentissage automatique : alléger les calculs de mécanique quantique des méthodes premier principe tout en amoindrissant la perte de précision associée. Pour accomplir cela, on effectue nombre d'approximations basées sur une forte intuition physique qui permettent de pré-calculer des paramètres clés, tout en assurant une perte de précision marginale. Si cette approche apporte un gain de temps inférieur à ses analogues complètement paramétriques, elle garantit de conserver l'accès aux propriétés électroniques du système considéré et assure une forte robustesse.

Dans le formalisme *DFTB*, on approxime le potentiel répulsif comme une somme de contributions polynomiales dépendantes de la distance entre deux atomes, aux paramètres ajustés pour coller aux résultats premier principe. Cette représentation est limitée par sa forme fonctionnelle simple, et il s'agit d'une simplification majeure de la méthode, avec une justification physique moindre. Les présents travaux se concentrent ainsi sur un potentiel complètement paramétré, ajusté à la main et à la faible motivation physique intrinsèque : en somme, un candidat tout désigné pour l'application de méthodes d'apprentissage automatique.

Le chapitre 2 de cette thèse présente le formalisme de la méthode premier principe de la théorie de la fonctionnelle de la densité (DFT), qui sert de base de développement à la méthode DFTB, détaillée au chapitre 3. Ceci nous permet d'énoncer clairement les objectifs de ces travaux au chapitre 4. Nous portons ensuite notre intérêt sur les méthodes d'apprentissage automatique au chapitre 5. Cela nous mène à la première partie des travaux de cette thèse dans le chapitre 6, qui détaille les formalismes en jeu et leur implémentation. Le chapitre 7 présente les potentiels répulsifs générés durant les présents travaux en deux points : l'acquisition de données de référence et l'entraînement de réseaux de neurones. Une évaluation extensive des apports de cette nouvelle approche pour une large variété de systèmes de Silicium pur est ensuite donnée au chapitre 8. Enfin, le chapitre 9 conclut en résumant les travaux effectués, discute les points majeurs relevés ainsi que plusieurs directions à approfondir ou explorer en perspective.

Chapitre 2

Potentiels premier principe

Dans ce chapitre, nous introduisons succinctement la théorie de la fonctionnelle de la densité (*Density Functional Theory* - **DFT**), de ses motivations aux possibilités qu'elle offre, en passant par son formalisme mais aussi quelques unes de ses principales approximations. Ce sujet étant fréquemment détaillé dans la littérature, cette section ne se veut pas exhaustive et s'applique à présenter les concepts nécessaires à la compréhension du cadre théorique dans lequel s'inscrit cette thèse.

Appréhender le comportement de la matière au niveau atomique implique de considérer des systèmes d'une taille caractéristique de l'ordre de quelques Angströms (1 Å = 10^{-10} m). À cette échelle, les lois de la mécanique quantique sont dominantes. Dans une approximation non-relativiste, déterminer les propriétés du système revient alors à résoudre une des équations fondamentales de la théorie quantique : *l'équation de Schrödinger*, que l'on introduit brièvement plus loin. Nous proposons de présenter des méthodes permettant de résoudre numériquement cette équation dans les quelques pages suivantes : en partant du formalisme le plus fondamental, nous justifions les principales approximations par étapes chronologiques menant à l'apparition de la méthode *DFT*.

2.1 Introduction historique

En décembre 1926, Erwin Schrödinger publie un article traitant de la théorie ondulatoire de la mécanique des atomes et molécules [1]. De manière hésitante, conscient du changement fondamental de perspective sur notre vision du monde qu'il apporte, il propose une solution visant à outrepasser les limitations de la mécanique classique (qu'il nomme "ordinary mechanics") à représenter les systèmes de très petite taille, c'est-à-dire à l'échelle atomique ou subatomique. Le formalisme proposé s'appuie sur la thèse soutenue par Louis de Broglie en 1924 [2], et implique, entre autres, d'attribuer aux particules du système un comportement ondulatoire.

Les travaux présentés par Schrödinger, dans cet article et d'autres, redéfinissent notre rapport aux petites échelles, donnant naissance dans les années suivantes à la **mécanique quantique**; de nombreuses recherches menées par la suite complètent cette théorie, jusqu'à en faire l'un des socles de la physique moderne.

La mécanique quantique postule notamment que l'état $|\Psi, t\rangle$ de tout système quantique non-relativiste est une solution de l'équation de Schrödinger dépendante du temps.

$$\hat{H}|\Psi,t\rangle = i\hbar \frac{\partial|\Psi,t\rangle}{\partial t}.$$
 (2.1)

Bien qu'il existe en principe une solution exacte à cette équation quel que soit le système considéré, il n'est dans les faits possible de la résoudre que pour des systèmes très simples : dès lors qu'il est nécessaire de considérer plusieurs particules en interaction, la solution, comme pour nombre d'autres équations différentielles, ne peut pas à notre connaissance être exprimée de manière analytique et exacte.

Pour pallier cette limitation fondamentale, plusieurs approximations et algorithmes ont été développés dans l'objectif de fournir des solutions numériques suffisamment précises pour comprendre les mécanismes en jeu aux petites échelles.

2.2 Théorie de Hartree-Fock

La représentation d'un système atomique ne dépendant pas explicitement du temps implique la résolution de l'équation de Schrödinger indépendante du temps :

$$\hat{H}|\Psi\rangle = E|\Psi\rangle,$$
 (2.2)

où \hat{H} est l'opérateur "énergie totale", ou hamiltonien, du système et la fonction d'onde $|\Psi\rangle$ un vecteur normalisable d'un espace de Hilbert¹ décrivant complètement l'état de toutes les particules du système, électrons comme noyaux atomiques.

L'objectif est d'obtenir les énergies E du système ainsi que les fonctions d'ondes associées desquelles découlent toutes ses propriétés.

Le hamiltionien du système est défini comme l'opérateur correspondant à l'énergie totale du système et est exprimé par analogie à la mécanique classique comme la somme de l'opérateur énergie cinétique \hat{T} et de l'opérateur énergie potentielle \hat{V} :

$$\hat{H} = \hat{T} + \hat{V},$$

que l'on peut ré-exprimer, en unités atomiques, pour un système à N_e électrons et N_n noyaux atomiques comme :

$$\hat{H} = -\sum_{i=1}^{N_e} \frac{1}{2} \nabla_i^2 - \sum_{I=1}^{N_n} \frac{1}{2M_I} \nabla_I^2 - \sum_{i=1}^{N_e} \frac{Z_I}{|\vec{r_i} - \vec{R_I}|} + \sum_{i=1}^{N_e} \sum_{j>i}^{N_e} \frac{1}{|\vec{r_i} - \vec{r_j}|} + \sum_{I=1}^{N_n} \sum_{J>I}^{N_n} \frac{Z_I Z_J}{|\vec{R_I} - \vec{R_J}|},$$
(2.3)

1. Espace qui étend les méthodes d'algèbre linéaire en généralisant les notions d'espaces euclidiens ou hermitiens à des espaces de dimensions quelconques (finies ou infinies).

où :

- $\vec{r_i}$ désigne la position de l'électron *i*.
- \hat{R}_I désigne la position du noyau atomique I.
- Z_I est le numéro atomique du noyau atomique I.
- M_I désigne la masse du noyau atomique I.

Par identification, la première ligne de l'équation (2.3) correspond à l'opérateur énergie cinétique $\hat{T} = \hat{T}_e + \hat{T}_n$. Les deux termes représentent respectivement les opérateurs énergie cinétique des électrons \hat{T}_e et des noyaux atomiques \hat{T}_n . Les opérateurs Laplaciens agissent sur les positions des électrons et des noyaux atomiques de la manière suivante :

$$\nabla_i^2 A = \frac{\partial^2 A}{\partial r_{i,x}^2} + \frac{\partial^2 A}{\partial r_{i,y}^2} + \frac{\partial^2 A}{\partial r_{i,z}^2} \quad \text{et} \quad \nabla_I^2 A = \frac{\partial^2 A}{\partial R_{I,x}^2} + \frac{\partial^2 A}{\partial R_{I,y}^2} + \frac{\partial^2 A}{\partial R_{I,z}^2},$$

en coordonnées cartésiennes, pour une fonction A de l'espace de Hilbert.

La seconde ligne de l'équation (2.3) correspond à l'opérateur énergie potentielle du système \hat{V} . Ces trois termes d'interaction coulombienne représentent respectivement :

- $\hat{V}_{n \leftrightarrow e}$ l'attraction entre noyaux et électrons.
- $V_{e \leftrightarrow e}$ la répulsion entre électrons.
- $\hat{V}_n \leftrightarrow_n$ la répulsion entre noyaux atomiques.

La fonction d'onde Ψ est définie comme une fonction de la position et des spins de toutes les particules du système.

$$\Psi = \Psi \left(\left\{ \vec{r}_i, \omega_i \right\}_{i=1}^{N_e}, \left\{ \vec{R}_I, \Omega_I \right\}_{I=1}^{N_n} \right).$$

Notre présentation ne dépend pas de manière importante des variables de spin ω_i et Ω_I , nous les omettons donc dans la suite. Par ailleurs nous nous plaçons dans le cadre de l'approximation de *Born-Oppenheimer* [3], en séparant les degrés de libertés électroniques et nucléiques et en considérant les noyaux comme des particules classiques. Dans cette approximation les fonctions d'ondes des électrons, légers et rapides, s'adaptent de manière instantanée aux positions des noyaux traités comme des charges ponctuelles positives. L'approximation de Born-Oppenheimer peut se justifier par le petit rapport de masse $\frac{m_e}{M_I}$, où m_e est la masse de l'électron.

Dans ce cadre, l'équation à résoudre devient celle donnée par le Hamiltonien d'un système de N_e électrons dans le potentiel créé par les noyaux :

$$\hat{H}_e = \sum_{i=1}^{N_e} \left[\frac{1}{2} \nabla_i^2 + \sum_{j>i}^{N_e} \frac{1}{|\vec{r_i} - \vec{r_j}|} - \sum_{I=1}^{N_n} \frac{Z_I}{|\vec{r_i} - \vec{R_I}|} \right].$$
(2.4)

Dans cette dernière équation, les coordonnées des noyaux représentent des paramètres fixés pour lesquels sont déterminées les fonctions d'ondes électroniques. Les fonctions d'ondes solutions de l'équation stationnaire de Schrödinger dépendent donc de manière paramétriques des positions atomiques :

$$\hat{H}_e(\{\vec{R}_I\}_{I=1}^{N_n})|\psi(\{\vec{R}_I\}_{I=1}^{N_n})\rangle = E_e(\{\vec{R}_I\}_{I=1}^{N_n})|\psi(\{\vec{R}_I\}_{I=1}^{N_n})\rangle.$$
(2.5)

Cette approche est intéressante : lorsqu'on considère l'énergie cinétique des noyaux ainsi que l'interaction entre les noyaux $V_{n \leftrightarrow n}(\{\vec{R}_I\}_{I=1}^{N_n})$, on peut reconstruire un Hamiltonien pour les noyaux.

$$H_n = T_n \left(\left\{ \vec{R}_I \right\}_{I=1}^{N_n} \right) + V_n \longleftrightarrow_n \left(\left\{ \vec{R}_I \right\}_{I=1}^{N_n} \right) + E_e \left(\left\{ \vec{R}_I \right\}_{I=1}^{N_n} \right).$$
(2.6)

Dans cette équation, E_e et $V_n \leftrightarrow n$ jouent le rôle de potentiels classiques agissant sur les coordonnées nucléaires, bien que le premier dépende directement de l'état quantique des électrons. On peut donc faire évoluer classiquement les positions des noyaux dans le champ quantique des électrons : ce sont les bases des méthodes de dynamique classique des noyaux, telles que la dynamique moléculaire de Born-Oppenheimer [4].

Revenons à la résolution du problème électronique. En supposant les ions fixés, on réécrit l'équation stationnaire (2.5) plus simplement :

$$\hat{H}_e|\psi\rangle = E_e|\psi\rangle. \tag{2.7}$$

En 1928, Douglas Hartree propose d'approcher la fonction d'onde $|\psi\rangle$ à N_e électrons par un produit de fonctions d'ondes mono-électroniques qui fut par la suite appelé *produit* de Hartree [5] :

$$\psi\left(\{\vec{r}_i\}_{i=1}^{N_e}\right) = \psi_1\left(\vec{r}_1\right) \cdot \psi_2\left(\vec{r}_2\right) \cdot \dots \cdot \psi_{N_e}\left(\vec{r}_{N_e}\right).$$
(2.8)

Bien qu'il représente un premier pas significatif pour exprimer la fonction d'onde à N_e électrons, le produit de Hartree ne possède pas de symétrie par rapport à l'échange de deux coordonnées électroniques : $\vec{r_1} \leftrightarrow \vec{r_2}$.

Une expression plus complète du principe de Pauli [6], prenant en compte l'indiscernabilité des fermions, impose que la fonction d'onde des électrons soit antisymétrique vis-à-vis de l'échange des coordonnées de deux particules (*de même spin, non précisé dans notre explication simplifiée*) :

$$\psi\left(\cdots,\vec{r_i},\cdots,\vec{r_j},\cdots\right) = -\psi\left(\cdots,\vec{r_j},\cdots,\vec{r_i},\cdots\right).$$

La méthode Hartree-Fock respecte par construction cette règle de symétrie en recherchant d'emblée une solution antisymétrique sous la forme d'un *déterminant de Slater*, proposé en 1929 par John Clarke Slater [7] :

$$\psi\left(\{\vec{r}_{i}\}_{i=1}^{N_{e}}\right) = \frac{1}{\sqrt{N_{e}!}} \begin{bmatrix} \psi_{1}\left(\vec{r}_{1}\right) & \psi_{2}\left(\vec{r}_{1}\right) & \cdots & \psi_{N_{e}}\left(\vec{r}_{1}\right) \\ \psi_{1}\left(\vec{r}_{2}\right) & \psi_{2}\left(\vec{r}_{2}\right) & \cdots & \psi_{N_{e}}\left(\vec{r}_{2}\right) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{1}\left(\vec{r}_{N_{e}}\right) & \psi_{2}\left(\vec{r}_{N_{e}}\right) & \cdots & \psi_{N_{e}}\left(\vec{r}_{N_{e}}\right) . \end{bmatrix}$$
(2.9)

La résolution des approches de Hartree et de Hartree-Fock procède grâce au principe variationnel : les variations de l'énergie autour des énergies propres du Hamiltonien électronique sont stationnaires lorsque les fonction d'ondes correspondent à des états propres.

$$\delta \frac{\langle \psi | \hat{H} | \psi \rangle}{\langle \psi | \psi \rangle} = 0.$$
(2.10)

L'application du principe variationnel au Hamiltonien électronique, en utilisant un déterminant de Slater et les conditions d'orthonormalisation des fonctions d'ondes, mène à l'équation de Fock qui est une équation de Schrödinger indépendante du temps appliquée au problème électronique.

$$\left[-\frac{\nabla^2}{2} + V_0(\vec{r}) + V_H(\vec{r})\right]\psi_k(\vec{r}) - \sum_{l=1}^{N_e}\psi_l(\vec{r})\int d^3r' \frac{\psi_k(\vec{r'})\psi_l^*(\vec{r'})}{\left|\vec{r} - \vec{r'}\right|} = \varepsilon_k\psi_k(\vec{r}), \quad (2.11)$$

où $V_0(\vec{r})$ est le potentiel d'interaction d'un électron avec les noyaux atomiques

$$V_0(\vec{r}) = \sum_{I=1}^{N_n} \frac{Z_I}{\left| \vec{r} - \vec{R}_I \right|},$$

et $V_H(\vec{r})$ est un potentiel de type Hartree, qui représente l'interaction Coulombienne d'un électron avec la densité de charge $\rho(\vec{r}) = \sum_{l=1}^{N_e} \psi_l^*(\vec{r}) \psi_l(\vec{r})$:

$$V_H(\vec{r_i}) = \sum_{l=1}^{N_e} \int d^3r \frac{\psi_l^*(\vec{r})\psi_l(\vec{r})}{|\vec{r} - \vec{r_i}|} = \int d^3r \frac{\rho(\vec{r})}{|\vec{r} - \vec{r_i}|}.$$

Le dernier terme du membre de gauche est *le potentiel d'échange* appliqué à la fonction d'onde $|\psi_k\rangle$; il est plus complexe que les autres. Le terme d'échange est négatif, non-local et dépend du chevauchement des fonctions d'ondes du système. De plus *(même si cela n'est pas apparent dans notre présentation simplifiée)*, il ne couple que des fonctions d'ondes $\psi_k(\vec{r})$ et $\psi_l(\vec{r})$ de même spin. Ce terme d'échange découle directement de la forme antisymétrique du déterminant de Slater et donc du principe de Pauli. Bien qu'il soit issu d'une théorie quantique qui s'applique à très petite échelle, il est à l'origine des phénomènes de ferromagnétisme et d'anti-ferromagnétisme qui s'expriment à l'échelle macroscopique dans les métaux de transition.

Sans entrer dans le détail de la résolution du problème de Hartree-Fock, notons certaines caractéristiques importantes du traitement de l'équation (2.11) que l'on retrouve dans d'autres méthodes de structure électronique.

Insistons tout d'abord sur le fait que la théorie de Hartree-Fock est une théorie de l'état fondamental. En effet, le principe variationnel revient en pratique à minimiser l'énergie du système d'électrons, permettant de déterminer l'état fondamental du système alors que les états excités restent en général hors de portée d'une telle procédure. D'autre part, on remarque que le potentiel utilisé dans cette méthode variationnelle dépend lui même des solutions $\psi_k(\vec{r})$. Cette interdépendance nécessite donc une résolution itérative et auto-cohérente (*Self Consistent Field Method*). En pratique, ceci rend la tâche plutôt difficile et coûteuse numériquement, en particulier lorsqu'un grand nombre d'électrons sont impliqués.

La résolution de l'équation (2.11) utilise une base de fonctions d'ondes sur laquelle on exprime les différents termes d'énergie et décompose les fonctions mono-électroniques $|\psi_k\rangle$.

Les solutions auto-cohérentes $|\psi_k\rangle$ représentent un jeu d'orbitales $\{\psi_k\}$ orthonormales d'énergies $\{\varepsilon_k\}$. Le déterminant de Slater obtenu à partir des N_e orbitales de plus basses énergies est l'état fondamental de Hartree-Fock ψ_0 avec l'énergie totale E_0^{HF} .

Le sens physique des énergies d'orbitales $\{\varepsilon_k\}$ nous est donné par le théorème de Koopmans, démontré en 1934 par Tjalling Koopmans [8], qui recevra par ailleurs le prix Nobel d'économie en 1975 :

L'énergie ε_k associée à une orbitale occupée correspond, au signe près, à l'énergie d'ionisation nécessaire pour arracher l'électron qui occupe cette orbitale ψ_k .

Théorème de Koopmans

En d'autres termes, il s'agit d'une approximation dans laquelle on néglige l'impact sur les autres orbitales de l'ajout ou du retrait d'un électron dans une orbitale. Cela signifie que l'on ne prend pas en compte la relaxation des orbitales induite par de tels changements, et on parle parfois d'*approximation des orbitales gelées*. Cette approche peut faire office d'une première approximation acceptable pour le calcul de potentiels d'ionisation, mais la mauvaise représentation des affinités électroniques cause une tendance à surévaluer les gaps et à sous-évaluer la largeur des bandes lors des calculs de structure électronique.

Dans la méthode Hartree-Fock, l'énergie totale $E_0^{\rm HF}$ est obtenue par l'application du principe variationnel en supposant que la fonction d'onde prend la forme d'un déterminant de Slater. Le terme d'échange est une contribution majeure aux corrélations effectives entre électrons de même spin, mais la méthode Hartree-Fock ne décrit pas précisément les corrélations entre électrons de spins opposés. Si on note E_0 l'énergie correcte du système de N_e électrons, les contraintes sur la forme de la fonction d'onde Hartree-Fock imposent donc :

$$E_0 < E_0^{\mathrm{HF}}$$

Bien qu'il s'agisse effectivement d'une corrélation entre électrons de même spin, on l'appelle simplement *"échange"* dans le vocabulaire de la structure électronique. En revanche, on nomme *"corrélation"* les termes correctifs au-delà de Hartree-Fock. Ceci définit l'énergie de corrélation comme une quantité négative :

$$E_c = E_0 - E_0^{\rm HF} < 0.$$

Pour aller au-delà de cette limite imposée par la forme de la fonction d'onde, on peut considérer un ensemble de déterminants de Slater initialement construits en occupant des orbitales mono-électroniques de plus haute énergie. On fait référence à ces méthodes sous le nom de méthodes "post Hartree-Fock". Néanmoins, malgré des efforts récents [9], la scalabilité de la méthode Hartree-Fock reste plutôt mauvaise (originellement $O(N^4)$ où N est le nombre de fonctions de base avec $N \ge N_e$ [10]). Les méthodes post Hartree-Fock, nécessairement plus lourdes numériquement, se limitent en pratique à des systèmes de quelques dizaines d'atomes. Une alternative méthodologique est donnée par la théorie de la fonctionnelle de la densité dans laquelle on peut tenir compte à la fois des termes d'échange et de corrélation.

2.3 Théorie de la fonctionnelle de la densité

En 1964, Pierre Hohenberg et Walter Kohn proposent d'approcher le problème de la structure électronique sous un nouvel angle [11]. L'idée originale s'appuie sur le modèle de Thomas-Fermi, proposé en 1927 par Llewellyn Thomas [12] et Enrico Fermi [13], et montre que l'énergie totale d'un ensemble d'électrons en interaction dans un potentiel externe peut être écrite comme une fonctionnelle de la densité électronique $E[n(\vec{r})]$, au lieu de la fonction d'onde à N_e électrons $\psi\left(\{\vec{r}_i\}_{i=1}^{N_e}\right)$:

Toute observable physique du système à l'état fondamental est également une fonctionnelle unique et universelle de la densité électronique n_0 , soit pour l'énergie : $E = E[n_0(\vec{r})].$

Premier théorème de Hohenberg-Kohn

Un an plus tard, W. Kohn et Lu Sham proposent un nouvel algorithme permettant d'obtenir d'une part la densité électronique n et d'autre part d'en déduire l'énergie totale associée E[n] [14]. On exprime la fonctionnelle énergie totale comme

$$E[n] = T_0[n] + E_{\rm H}[n] + E_{\rm xc}[n] + \int V_{\rm ext}(\vec{r})n(\vec{r})d^3r, \qquad (2.12)$$

où

- $T_0[n]$ est la fonctionnelle énergie cinétique correspondant à l'énergie cinétique d'un gaz inhomogène de densité n composé d'électrons n'interagissant pas.
- $E_{\rm H}[n]$ est l'énergie de Hartree (répulsion électron-électron).

$$E_{\rm H}[n] = \frac{1}{2} \int n(\vec{r}) V_{\rm H}(\vec{r}) d^3r = \frac{1}{2} \int \int n(\vec{r}) \frac{n(\vec{r})}{\left|\vec{r} - \vec{r'}\right|} d^3r d^3r'.$$

- $E_{\rm xc}[n]$ est l'énergie d'échange-corrélation. C'est la somme de l'énergie d'échange E_x , de l'énergie de corrélation E_c et d'une correction sur l'énergie cinétique provenant des corrélations entre électrons. Sa dépendance exacte en n n'est pas connue et doit être approximée.
- $V_{\text{ext}}(\vec{r})$ est un potentiel externe agissant sur le système électronique (par exemple les charges des noyaux atomiques).

La forme fonctionnelle de l'énergie cinétique d'un gaz d'électrons en interaction en fonction de la densité étant inconnue, Kohn et Sham choisissent une nouvelle approche.

Si vous n'aimez pas la réponse, modifiez la question.

Ansatz de Kohn et Sham

Ils proposent de représenter la densité n comme la somme d'orbitales à un électron ψ_i d'occupations respectives f_i . Pour M orbitales prises en compte dans le calcul, on a :

$$n(\vec{r}) = \sum_{i=1}^{M} f_i |\psi_i(\vec{r})|^2, \quad \text{où } \sum_{i=1}^{M} f_i = N_e.$$
(2.13)

On peut donc exprimer l'énergie en fonction de ces orbitales à un électron.

$$E[n] = \sum_{i=1}^{M} f_i \int \psi_i^*(\vec{r}) \left(-\frac{\nabla^2}{2} + V_{\text{ext}}(\vec{r}) + \frac{1}{2} V_{\text{H}}[n(\vec{r})] \right) \psi_i(\vec{r}) d^3r + E_{\text{xc}}[n].$$

A partir de cette expression, le second théorème de Hohenberg-Kohn fournit un principe variationnel permettant de déterminer l'état fondamental du système à partir de la fonctionnelle E[n].

Pour un potentiel et un nombre d'électrons donnés, l'énergie de l'état fondamental du système est la valeur qui minimise la fonctionnelle E[n], et la densité qui lui est associée correspond à la densité exacte $n_0(\vec{r})$ de l'état fondamental.

Second théorème de Hohenberg-Kohn

On a donc

$$\frac{\delta E[n(\vec{r})]}{\delta n(\vec{r})}\bigg|_{n_0(\vec{r})} = 0,$$

d'où, pour l'état fondamental

$$\frac{\delta}{\delta\psi_i^*(\vec{r})} \left[E\left[\left\{\psi_j^*\right\}_{j=1}^M\right] - \sum_{j,k=1}^M \lambda_{jk} \int \psi_j^*(\vec{r'})\psi_k(\vec{r'})d^3r' \right] = 0 , \forall i,$$

où les multiplicateurs de Lagrange λ_{jk} permettent de tenir compte de l'orthonormalisation des orbitales ψ_j . On peut alors écrire les équations de Kohn-Sham :

$$\left[-\frac{\nabla^2}{2} + V_{\text{ext}}(\vec{r}) + V_{\text{H}}[n(\vec{r})] + V_{\text{xc}}[n(\vec{r})]\right]\psi_i(\vec{r}) = \varepsilon_i\psi_i, \qquad (2.14)$$

où le nouveau potentiel

$$V_{\rm xc}[n(\vec{r})] = \frac{\delta E_{\rm xc}[n(\vec{r})]}{\delta n(\vec{r})}$$

désigne la dérivée fonctionnelle de l'énergie d'échange-corrélation.

Nous avons vu plus tôt (2.13) que la densité $n(\vec{r})$ dépend des orbitales $\psi_i(\vec{r})$. L'opérateur agissant sur ces orbitales dépend donc indirectement de ces mêmes orbitales. De manière analogue à la théorie de Hartree-Fock (2.10), il faut recourir à un algorithme auto-cohérent (*Self Consistent Field Method, SCF*) pour résoudre les équations de Kohn-Sham (2.14). À partir d'une proposition acceptable pour densité initiale, on obtient des orbitales associées nous permettant de déterminer une nouvelle densité, donc de nouvelles orbitales.. Lorsque la différence entre deux calculs successifs devient négligeable, on considère que le système d'équation a convergé et on arrête la procédure.

L'énergie totale DFT peut s'écrire comme :

$$E[n(\vec{r})] = \sum_{i}^{M} f_{i}\varepsilon_{i} - E_{\rm H}[n(\vec{r})] - \int n(\vec{r})V_{\rm xc}[n(\vec{r})]d^{3}r + E_{\rm xc}[n(\vec{r})]$$

$$= -\frac{1}{2}\sum_{i}^{M} f_{i}\int\psi_{i}^{*}(\vec{r})\nabla^{2}\psi_{i}(\vec{r}) + \int n(\vec{r})V_{\rm ext}[n(\vec{r})]d^{3}r + E_{\rm H}[n(\vec{r})] + E_{\rm xc}[n(\vec{r})].$$

(2.15)

2.3.1 Approximation de la densité locale

L'énergie d'échange et corrélation $E_{\rm xc}$ est inconnue et doit être approximée. Parmi pléthore d'approches développées au cours des dernières décennies, nous proposons de présenter la classe la plus simple d'entre elles : l'approximation de la densité locale (*Local Density Approximation - LDA*).

Dans cette approximation, la fonctionnelle d'échange et corrélation s'écrit comme une simple somme de contributions calculées en chaque point de l'espace :

$$E_{\rm xc}^{\rm LDA}\left[n(\vec{r})\right] = \int n(\vec{r})\varepsilon_{\rm xc}\left(n(\vec{r})\right)d^3r.$$
(2.16)

L'approche la plus courante consiste à approximer l'énergie d'échange et corrélation $\varepsilon_{\rm xc}(n(\vec{r}))$ par la densité électronique d'un gaz homogène d'électrons de densité $n(\vec{r})$. On décompose cette énergie comme la somme d'une énergie d'échange ε_x et d'une énergie de corrélation ε_c .

$$\varepsilon_{\mathrm{xc}}(n) = \varepsilon_{x}(n) + \varepsilon_{c}(n)$$

Pour un gaz homogène d'électrons, l'énergie d'échange est connue analytiquement :

$$\varepsilon_x(n) = -\frac{3}{4} \left(\frac{3}{\pi}\right)^{1/3} n^{1/3}.$$

Le calcul de l'énergie de corrélation est moins aisé. Des expressions analytiques pour un gaz homogène d'électrons sont connues dans les limites des hautes et basses densités, correspondant respectivement à des corrélations infiniment faibles ou infiniment fortes.

Plusieurs approches, numériques [15] ou non [16], sont disponibles pour permettre de déterminer l'énergie de corrélation pour toutes les densités.

Le potentiel d'échange et corrélation associé à la fonctionnelle (2.16) est :

$$V_{\rm xc}^{\rm LDA}\left[n(\vec{r})\right] = \frac{\delta E_{\rm xc}^{\rm LDA}\left[n(\vec{r})\right]}{\delta n(\vec{r})} = \varepsilon_{\rm xc}\left(n(\vec{r})\right) + \frac{d\varepsilon_{\rm xc}\left(n(\vec{r})\right)}{dn(\vec{r})}n(\vec{r}).$$
 (2.17)

Bien que l'approximation sur laquelle repose la fonctionnelle LDA puisse paraître simpliste, celle-ci est suprenamment performante pour de nombreux systèmes. En pratique, la LDA tend à sous-évaluer systématiquement $E_{\rm xc}$ pour les points de basse densité, et compense en surévaluant $E_{\rm xc}$ pour les points de densité plus forte. En moyenne, on obtient un comportement satisfaisant : les propriétés élastiques, structurelles et vibrationnelles de systèmes où la distribution de densité est relativement uniforme sont bien représentées par cette approximation.

En revanche, la méthode tend à favoriser les systèmes à forte densité électronique. La surévaluation de la cohésion des systèmes implique des énergies de liaisons et de formations élevées, mais aussi des constantes élastiques trop grandes et une sous-estimation des distances de liaisons [17]. Les liaisons hydrogènes [18] ou les interactions de type Van der Waals sont mal traitées. Enfin, l'approximation rencontre des difficultés à représenter les surfaces et plus généralement les systèmes où les gradients de densité sont forts.

2.3.2 Approximation du gradient généralisé

Pour s'affranchir des difficultés posées par le calcul de l'énergie de corrélation en LDA, l'approximation du gradient généralisé (*Generalized Gradient Approximation -* **GGA** [19]) propose de tenir compte, non seulement de la densité électronique $n(\vec{r})$, mais aussi de son gradient $\nabla n(\vec{r})$.

$$E_{\rm xc}^{\rm GGA}\left[n(\vec{r}), \nabla n(\vec{r})\right] = \int n(\vec{r}) \varepsilon_{\rm xc}\left(n(\vec{r}), \nabla n(\vec{r})\right) d^3r.$$
(2.18)

Bien que cette nouvelle fonctionnelle apporte une souplesse supérieure à la fonctionnelle *LDA*, ainsi qu'une meilleure représentation des énergies totales, cela s'accompagne d'une augmentation sensible du nombre de paramètres. Parmi les nombreuses paramétrisations GGA disponibles, dans le cadre des calculs DFT effectués lors de cette thèse nous avons choisi d'utiliser la paramétrisation Perdrew-Burke-Ernzerhof (**PBE**) [20] pour sa capacité à fournir des résultats robustes pour une large gamme de systèmes.

Bien que d'autres fonctionnelles plus sophistiquées existent (B3LYP [21], meta-GGA [22]), leur possible meilleure précision quantitative s'accompagne d'un surcoût numérique notable par rapport à PBE. Les présents travaux nécessitent un très grand nombre de calculs DFT (voir 7.1); ce comportement nous semble préjudiciable, et on préfère donc la fonctionnelle PBE.

Chapitre 3

La méthode *Density-Functional* based Tight-Binding

Les perspectives offertes par la théorie de la fonctionnelle de la densité sont notables. En rendant possible d'approximer la résolution de problèmes électroniques à plusieurs atomes, elle permet de calculer de nombreuses propriétés pour des systèmes variés : entre autres, structure de bandes électroniques d'un solide, propriétés élastiques d'un matériau ou d'une molécule. La disposition de paramètres précis et parfaitement contrôlés parfois inaccessibles expérimentalement en font un outil de choix pour orienter, compléter ou corriger l'expérience et ainsi parfaire notre connaissance de la matière à petite échelle.

Le coût numérique d'un calcul DFT est cependant un facteur limitant de manière déterminante la taille des systèmes considérés et les durées étudiées. Il est ultimement nécessaire de diagonaliser un hamiltonien de taille proportionnelle au nombre N de fonctions de base du système : le coût typique d'une telle opération étant d'ordre $O(N^3)^1$, le temps de calcul *DFT* augmente fortement avec la taille du système. Au-delà de quelques centaines d'atomes sur des durées d'au plus quelques picosecondes, les implémentations de la méthode atteignent leur limite et les calculs ne sont plus réalisables.

Dans l'usage commun en simulation atomistique, la représentation empirique des interactions entre atomes représente une alternative aux méthodes dites *ab-initio* ou premier principe (résolvant les équations de la mécanique quantique) telles que DFT. Les potentiels empiriques simplifient les interactions atomiques en les décrivant sous la forme de fonctions fixes avec quelques paramètres ajustables qui sont déterminés pour reproduire des comportements précis pour des systèmes donnés. Le coût de calcul est donc considérablement réduit, et l'on peut sans peine s'intéresser à des systèmes de l'ordre de millions d'atomes pour des durées de l'ordre de plusieurs nanosecondes. Là encore, cette approche présente ses limitations. Tout d'abord, la performance de tels potentiels est très dépendante des conditions pour lesquels ils ont été produits : quelle phase thermodynamique ? Pour quelle propriété ? De plus, ces potentiels ne fournissent aucune information sur les électrons du système. Il est donc impossible d'avoir connaissance de la structure électronique du système et des propriétés qu'elle détermine.

^{1.} Certains travaux aspirent à réduire ce coût numérique à O(N).



FIGURE 3.1 – Vue générale des méthodes pour représenter la matière et des échelles associées.

C'est dans ce contexte que s'inscrivent les méthodes semi-empiriques, permettant de conserver un accès à l'information électronique du système pour un coût numérique réduit via plusieurs approximations et la détermination préalable de grandeurs clés pour accélérer le calcul. La méthode *Density-Functional based Tight-Binding (DFTB)* est l'une d'elles et est présentée plus loin. En conservant la description de la structure électronique mais en recourant à quelques approximations supplémentaires, le coût de calcul est plus accessible. Ceci permet d'étendre la portée des méthodes premier principe à des systèmes plus grands, sur des durées plus longues.

Les modèles DFTB s'obtiennent à partir de la théorie de la fonctionnelle de la densité en développant la fonctionnelle d'énergie totale sur la densité électronique. Le détail des méthodes peut être trouvé dans les articles originaux [23–25], et celui de l'implémentation numérique est disponible dans [26]. On rappelle (2.15) que l'énergie totale DFT peut s'écrire :

$$E[n(\vec{r})] = \sum_{i}^{M} f_{i}\varepsilon_{i} - \frac{1}{2} \int \int n(\vec{r}) \frac{n(\vec{r})}{\left|\vec{r} - \vec{r'}\right|} d^{3}r d^{3}r' - \int n(\vec{r}) V_{\rm xc}[n(\vec{r})] d^{3}r + E_{\rm xc}[n(\vec{r})].$$

L'idée est de partir d'une densité de référence $n_0(\vec{r})$ obtenue par superposition des densités électroniques des atomes isolés [27]. On peut décrire le système avec cette densité

de référence perturbée par des fluctuations de densité :

$$n(\vec{r}) = n_0(\vec{r}) + \delta n(\vec{r}).$$
 (3.1)

Cette expression de la densité électronique permet de développer la fonctionnelle d'énergie totale :

$$E^{\text{DFTB}}[n_0(\vec{r}) + \delta n(\vec{r})] = E^0[n_0(\vec{r})] + E^1[n_0(\vec{r}), \delta n(\vec{r})] + E^2[n_0(\vec{r}), \delta n^2(\vec{r})] + E^3[n_0(\vec{r}), \delta n^3(\vec{r})] + \cdots .$$
(3.2)

L'ordre auquel on interrompt le développement de la fonctionnelle de l'énergie totale détermine le modèle de calcul utilisé : arrêter le développement au premier ordre équivaut à un calcul DFTB non auto-cohérent (*non-SCC DFTB*, également appelé **DFTB1** [28]); le terminer au deuxième ordre équivaut à un calcul DFTB auto-cohérent (*SCC DFTB* ou **DFTB2** [25]); aller jusqu'au troisième ordre permet de mieux traiter les systèmes chargés (**DFTB3** [29]).

On se propose de présenter successivement le formalisme sur lequel reposent les deux premiers ordres de ce modèle, utilisés lors de cette thèse.

3.1 Développement au premier ordre

Le modèle DFTB1 repose sur le développement de la fonctionnelle de l'énergie totale (3.2) au premier ordre.

$$E^{\text{DFTB1}}\left[n_0(\vec{r}) + \delta n(\vec{r})\right] = E^0\left[n_0(\vec{r})\right] + E^1\left[n_0(\vec{r}), \delta n(\vec{r})\right].$$
(3.3)

Le premier terme E_0 est discuté extensivement plus loin (voir 3.3). Le second terme correspond à l'énergie de structure de bande E^{BS} pour le système considéré :

$$E^{\rm BS} = E^1 \left[n_0(\vec{r}), \delta n(\vec{r}) \right] = \sum_{i=1}^{N_e} f_i \langle \psi_i | \hat{H^0} \left[n_0(\vec{r}) \right] | \psi_i \rangle, \tag{3.4}$$

où f_i correspond à l'occupation de ψ_i .

Formellement, le modèle repose sur la représentation des orbitales de Kohn-Sham comme une combinaison linéaire d'orbitales atomiques (Linear Combination of Atomic Orbitals - LCAO):

$$\psi_i(\vec{r}) = \sum_{\mu=1}^{N_b} c_{i\mu} \chi_\mu(\vec{r}), \qquad (3.5)$$

où n_b est le nombre de fonctions de bases χ_{μ} , et $c_{i\mu}$ sont les coefficients de la combinaison linéaire, que l'on doit déterminer.

Le jeu de fonctions de base $\{\chi_{\mu}\}$ est minimal, au sens où l'on utilise uniquement une fonction de base pour chaque état de moment angulaire occupé. Plus simplement, et comme on ne considère ici que les électrons de valence - les interactions entre électrons de cœur sont inclues dans le terme d'ordre zéro (voir 3.3) - on a donc : une fonction par orbitale *s*, trois fonctions par orbitale *p*, cinq fonctions par orbitale *d*, etc. Ce choix d'une base minimale rappelle le formalisme empirique des **liaisons fortes** (*tight-binding theory*), expliquant au passage le nom de la méthode *Density Functional based* **Tight-Binding**. Il est à noter que le calcul n'est pas possible en utilisant des orbitales atomiques pures comme fonctions de base, qui sont trop étendues dans l'espace. À la place, on recourt à des orbitales localisées sur les atomes, confinées par un potentiel extérieur $\left(\frac{r}{r_0}\right)^2$, et on résout une nouvelle équation de Kohn-Sham.

$$\left[\frac{-\nabla^2}{2} + v^{\text{eff}} \left[n_{\text{atome}}\right] + \left(\frac{r}{r_0}\right)^2\right] \chi_\mu = \varepsilon_\mu \chi_\mu.$$
(3.6)

Ici, le potentiel effectif v^{eff} de l'équation de Kohn-Sham correspond à

$$v^{\text{eff}}[n_{\text{atome}}] = V_{\text{ext}}(r) + V_H[n_{\text{atome}}] + V_{xc}[n_{\text{atome}}].$$

La résolution de ces équations fournit des orbitales quasi-atomiques légèrement comprimées utilisées pour décrire la densité électronique en situation de liaison. Le rayon de confinement pour les bases d'orbitales atomiques r_0 est habituellement détaillé dans les publications présentant les paramétrisations.

La base $\{\chi_{\mu}\}$ nous amène à un hamiltonien qui s'écrit alors $H^{0}_{\mu\nu} = \langle \chi_{\mu} | \hat{H}^{0} [n_{0}(\vec{r})] | \chi_{\nu} \rangle$; ses termes diagonaux $H^{0}_{\mu\mu}$ sont simplement donnés par les valeurs propres ε_{μ} de (3.6) [30]. La détermination des termes non diagonaux $H^{0}_{\mu\nu}$, s'écrit comme :

$$H^{0}_{\mu\nu} = \langle \chi_{\mu} | \hat{H}^{0} [n_{0}(\vec{r})] | \chi_{\nu} \rangle \approx \langle \chi_{\mu} | - \frac{\nabla^{2}}{2} + V_{AB} [n_{0}] | \chi_{\nu} \rangle , \ \mu \in A \ , \ \nu \in B,$$
(3.7)

où deux approches sont possibles pour exprimer V_{AB} ; soit on superpose les densités atomiques soit les potentiels effectifs monoatomiques [31] :

$$V_{AB}[n_0] = \begin{cases} v^{\text{eff}}[n_A + n_B] & \text{superposition des densités,} \\ v^{\text{eff}}[n_A] + v^{\text{eff}}[n_B] & \text{superposition des potentiels.} \end{cases}$$

On construit le hamiltonien pour une paire d'éléments A-B à partir des densités individuelles n_A et n_B . On obtient donc une matrice dont les éléments sont évalués numériquement pour chaque couple A-B; ceux-ci sont pré-calculés une seule fois, et sont simplement fonctions de la distance inter-atomique. Pour cela, on utilise la règle de *Slater-Koster* [32], tenant compte de l'orientation des orbitales considérés au sein du système.

La détermination des coefficients $c_{i\mu}$ se fait en résolvant les équations de Kohn-Sham pour *DFTB*. Comme celles-ci dépendent de tous les termes du développement sur la densité, leur forme exacte est présentée plus loin (3.17). On obtient une expression de l'énergie de structure de bandes E^{BS} qui dépend uniquement de la densité de référence n_0 .

$$E^{1}[n_{0}(\vec{r}), \delta n(\vec{r})] = E^{\text{BS}}[n_{0}(\vec{r})] = \sum_{i=1}^{N_{e}} \sum_{AB} \sum_{\mu \in A} \sum_{\nu \in B} f_{i}c_{i\mu}c_{i\nu}H^{0}_{\mu\nu}.$$
 (3.8)

3.2 Développement au second ordre

Si l'on développe la fonctionnelle d'énergie totale (3.2) au second ordre, il est nécessaire de calculer :

$$E^{2}\left[n_{0}(\vec{r}), \delta n^{2}(\vec{r})\right] = \frac{1}{2} \int \int \left(\frac{1}{\left|\vec{r} - \vec{r'}\right|} + \frac{\delta^{2} E_{\rm xc}\left[n(\vec{r})\right]}{\delta n(\vec{r}) \delta n(\vec{r'})}\Big|_{n_{0}}\right) \delta n(\vec{r}) \delta n(\vec{r'}) d\vec{r} d\vec{r'}.$$
 (3.9)

Dans la limite des grandes distances interatomiques, les contributions d'échange et corrélation disparaissent dans le formalisme LDA, et l'on peut considérer E^2 comme un terme d'interaction purement coulombien proportionnel à $1/|\vec{r} - \vec{r'}|$; on fait parfois référence à E^2 dans la littérature avec la notation E^{coul} .

On approxime la variation de densité $\delta n(\vec{r})$ en un point de l'espace comme la superposition de densités électroniques atomiques en ce point.

$$\delta n\left(\vec{r}\right) = \sum_{A} \delta n_A\left(\vec{r}\right).$$

On assume que ces contributions atomiques à la densité électronique sont sphériques, de largeur τ , centrées sur les positions des noyaux et que leur intensité décroît exponentiellement [25].

$$\delta n_A \left(\vec{r} \right) \approx \frac{1}{\sqrt{4\pi}} \left(\frac{\tau_A^3}{8\pi} e^{-\tau_A \left| \vec{r} - \vec{R_A} \right|} \right) \Delta q_A. \tag{3.10}$$

En négligeant dans un premier temps les contributions d'échange-corrélation et pour des largeurs de densités électroniques sphériques τ pour chaque atome du système discutées plus loin, le terme d'énergie E^2 peut être écrit [26] tel que :

$$E_{\vec{\tau}}^2 \left[n_0(\vec{r}), \delta n^2(\vec{r}) \right] = \frac{1}{2} \sum_{AB} \gamma_{AB} \left(\tau_A, \tau_B, R_{AB} \right) \Delta q_A \Delta q_B, \tag{3.11}$$

où R_{AB} est la distance entre les atomes et γ_{AB} est une fonction correspondant à l'intégrale spatiale normalisée du produit de deux densités électroniques n_A et n_B .

$$\gamma_{AB}\left(\tau_{A},\tau_{B},R_{AB}\right) = \int \int \frac{\delta n_{A}(\vec{r})\delta n_{B}(\vec{r'})}{\left|\vec{r}-\vec{r'}\right|} d\vec{r} d\vec{r'}.$$

L'énergie (3.11) dépend des charges de Mulliken $\{q_A\}$, où les fluctuations de charge atomiques $\Delta q_A = q_A - Z_A$ sont par rapport à la charge de l'atome neutre. Ainsi, le développement au second ordre permet au formalisme de tenir compte de la redistribution de charge induite sur les atomes par les fluctuations de densités électroniques. Ces charges sont dépendantes des coefficients d'orbitales moléculaires $c_{i\mu}$. Les équations Kohn-Sham de *DFTB*, présentées plus loin (3.17), doivent donc être résolues de manière auto-cohérente.

Dans le cas où les distances interatomiques sont plus petites et où les charges sont localisées sur un seul et même atome, l'évaluation de γ_{AA} nécessite en principe de connaître exactement la distribution de charges. Cette solution requiert un développement de la densité de charge sur une base d'orbitales souhaitablement construite; un tel calcul s'avère très coûteux, et l'on préfère chercher à l'approcher habilement. Comme remarqué par Pariser dès 1956 [33], γ_{AA} s'approxime comme la différence du potentiel d'ionisation atomique et de l'affinité électronique. Dès lors, γ_{AB} ne dépend que de la distance interatomique et de paramètres U_A et U_B , nommés paramètres de Hubbard. Ceux-ci sont obtensibles à partir de calculs DFT; il s'agit de la dérivée seconde de l'énergie d'un atome isolé par rapport au nombre d'occupation de l'orbitale occupée de plus haute énergie [25]. Cette approximation est particulièrement élégante, dans le sens où les constantes U_A obtenues ne sont pas empiriques, mais directement issues de calculs premier principe, et ne nécessitent aucun ajustement. Par ailleurs, ceux-ci ont un sens physique fort puisqu'ils sont directement liés à la dureté chimique. Celle-ci est inversement proportionnelle à la largeur τ de la densité de charges atomiques, ce qui est relativement intuitif : les anions (atomes diffusifs) ont une dureté chimique réduite.

Dans le formalisme DFTB2, il est cependant à noter que le paramètre de Hubbard est indépendant de l'état de charge du système, et est donc constant. En conséquence, il est par exemple impossible pour un anion d'avoir une dureté chimique différente de celle de l'atome neutre ou du cation; ce problème est notamment discuté dans [34]. Il est résolu par le développement au troisième ordre, non utilisé dans cette thèse, et dont le détail peut être trouvé dans [29].

3.3 Énergie répulsive

Nous nous intéressons ici au terme de l'ordre zéro du développement : l'énergie répulsive \mathbf{E}_{rep} , qui est le point focal des présents travaux. Par définition, $E^0[n_0(\vec{r})]$ est complètement indépendant de $\delta n(\vec{r})$ - autrement dit, de tout transfert de charge - et ne nécessite que de connaître la densité de référence n_0 . Il peut donc être déterminé pour un système de référence, et appliqué à de nouveaux systèmes : les différences majeures liées aux spécificités des systèmes sont en principe représentées par les autres contributions à l'énergie totale. Il s'agit là d'une caractéristique clé de ce terme, assurant une transférabilité maximale aux potentiels répulsifs. Ce terme s'exprime *rigoureusement* comme :

$$E^{0}[n_{0}(\vec{r})] = \frac{1}{2} \sum_{AB} \frac{Z_{A}Z_{B}}{r_{AB}} - \frac{1}{2} \int \int \frac{n_{0}(\vec{r})n_{0}(\vec{r'})}{\left|\vec{r} - \vec{r'}\right|} d\vec{r} d\vec{r'} - \int V_{\rm xc}\left[n_{0}(\vec{r})\right] n_{0}(\vec{r})d\vec{r} + E_{\rm xc}\left[n_{0}(\vec{r})\right].$$
(3.12)

En l'état, cette notation requiert notamment de déterminer les contributions d'échange et corrélation associées au système étudié. Nous avons vu précédemment (2.3.1 et 2.3.2) que leur expression analytique est inconnue et doit être approximée. Si elles sont notamment traités par le hamiltonien \hat{H}_0 dans le cas du développement au premier ordre, on adopte ici un choix bien plus radical. On considère ce terme comme une somme de potentiels répulsifs de paire V_{rep} , ce qui simplifie énormément le calcul.

$$E^{0}[n_{0}(\vec{r})] \approx E_{\text{rep}} = \frac{1}{2} \sum_{AB} V_{AB}^{\text{rep}}(r_{AB}).$$
 (3.13)

Usuellement [25], ces potentiels de paire adoptent la forme de polynômes ou de splines ² dont on ajuste les paramètres pour reproduire des données DFT [24]. L'idée est la suivante : pour une paire d'atomes A et B dont on fait varier la distance r_{AB} , on calcule la différence entre l'énergie de référence $DFT E_{DFT}(r_{AB})$ et l'énergie DFTB sans interaction répulsive associée.

Dans le cas d'un développement au premier ordre, cela revient à soustraire l'énergie de structure de bandes $E_{\text{DFTB}}^{\text{BS}}(r_{AB})$ (3.4).

$$E_{\rm rep}(r_{AB}) = E_{\rm DFT}(r_{AB}) - E_{\rm DFTB}^{\rm BS}(r_{AB}).$$
 (3.14)

Évidemment, dans le cas où l'on considère un développement de l'énergie totale au second ordre, le calcul de l'énergie répulsive (3.14) est à redéfinir.

$$E_{\rm rep}(r_{AB}) = E_{\rm DFT}(r_{AB}) - E_{\rm DFTB}^{\rm BS}(r_{AB}) - E_{\rm DFTB}^{\rm coul}(r_{AB}).$$
(3.15)

On peut trouver un exemple d'application de ce processus pour l'interaction répulsive entre atomes de carbone dans la figure 3.2, issue de [35]. En fonction de la distance interatomique, la nature de liaison entre atomes change : une liaison triple aux très courtes distances, et à mesure que l'on écarte les atomes la liaison devient double, puis simple. Dans chacun des cas, le choix est fait d'ajouter le nombre adéquat d'hydrogènes pour permettre aux atomes de carbone de former quatre liaisons au total.

Pour chacune de ces distances, on détermine conjointement l'énergie DFT et l'énergie DFTB sans interaction répulsive, puis on déduit leur différence $E_{\rm rep}$; ces énergies tracées sont ici translatées de manière à ce que les interactions répulsives s'annulent dans le cas d'atomes suffisamment éloignés.

Plus rigoureusement, on optimise un potentiel répulsif pour reproduire la différence d'énergie par liaison. On présente habituellement le calcul pour un dimère, mais il est

^{2.} Fonctions définies par morceaux par des polynômes.


FIGURE 3.2 – Calcul d'un potentiel répulsif de paire pour la liaison C-C [35].

tout à fait envisageable d'utiliser d'autres systèmes de référence. On peut par exemple s'intéresser à la thèse de A. Sieck [36], de laquelle provient la paramétrisation pour le Silicium utilisée dans les présents travaux. Dans ce cas précis, le potentiel répulsif est optimisé pour reproduire l'énergie par liaison des phases cubique simple et diamant du Silicium. Ainsi, on utilise plusieurs boîtes de simulation avec ces symétries, et l'on fait varier les distances en modifiant l'échelle de la boîte de simulation. On veut usuellement se restreindre aux contributions des premiers voisins uniquement; pour ces systèmes, on choisit donc le rayon de coupure $r_{\rm cut} = 2.54$ Å.

Si le potentiel obtenu est supposé transférable aux systèmes où la coordinence n'excède pas six, on note cependant une légère mésestimation des interactions répulsives pour des systèmes éloignés des références utilisées pour sa paramétrisation, par exemple pour le nano-cluster Si₅ (Fig. 3.3, issue de [36]).

Cette limitation peut s'expliquer par la difficulté à représenter exactement les interactions répulsives par un simple potentiel de paire. C'est en effet le constat effectué par certains travaux récents [37] visant par ailleurs à améliorer les potentiels répulsifs de paire en imposant par construction des contraintes, afin notamment d'assurer une fonction à deux corps strictement convexe. Si le gain en qualité de représentation associé est réel, on note cependant que la transférabilité demeure limitée : l'obtention d'un unique potentiel



FIGURE 3.3 – Évaluation du potentiel répulsif de paire pour le Silicium utilisé dans cette thèse, issu de [36].

de paire adapté à tous les systèmes semble hors d'atteinte, et l'on préconise d'aller au-delà de cette représentation simplifiée vers des interactions à plusieurs corps.

Quoi qu'il en soit, cette paramétrisation semble être une approximation raisonnable pour des distances interatomiques supérieures à 2.2 Å. Notons que la plage de distances couverte par ce potentiel ($r_{AB} \in [2.20, 2.54]$ Å), bien que contenant la majorité des distances importantes pour les phases du Silicium, est plutôt étroite. La borne inférieure de cet intervalle semble relativement élevée : pour quelques systèmes de Silicium, tels que les phases désorganisés comme les liquides, l'apparition de distances interatomiques de l'ordre de 2 Å est probable. Dans l'idéal, on souhaiterait pouvoir efficacement couvrir des distances plus faibles - il s'agit de l'un des travaux effectués lors de cette thèse.

En revanche, il n'est pas souhaitable d'étendre trop la portée du potentiel en augmentant à volonté le rayon de coupure $r_{\rm cut}$, borne supérieure de cet intervalle. Intrinsèquement, un potentiel répulsif rend compte des interactions entre les cœurs durs du système, d'interactions entre électrons et est utilisé pour corriger plusieurs double-comptages et approximations inhérents à la méthode. Ces termes sont purement liés à l'interaction entre atomes proches, et le comportement de monomères est supposément décrit par les contributions électroniques à l'énergie : $E_{\rm rep}$ est donc un terme courte-portée. Ainsi augmenter $r_{\rm cut}$ signifie tenir compte de différences entre DFT et les approximations DFTB des autres termes aux longues portées. Cela aide, en théorie, à reproduire plus efficacement les propriétés des calculs premier principe en tenant compte d'informations configurationnelles supplémentaires, mais induit une limitation majeure. Ainsi, inclure des corrections longue-portée aux approximations électroniques de DFTB dans l'énergie répulsive la rend, implicitement, dépendante des variations de charges Δq - et donc des fluctuations de densités δn . Dès lors, la transférabilité du potentiel, depuis le(s) système(s) de référence vers d'autres, n'est plus garantie. Par ailleurs, notons également que DFT et DFTB ne prédisent pas nécessairement les conversions intersystèmes aux mêmes distances; en conséquence, augmenter la portée du potentiel signifie reproduire la différence entre les deux méthodes, alors même que les deux peuvent ne pas décrire le même état électronique. En contraignant le potentiel répulsif aux courtes distances, on évite cet écueil.

En conséquence, il est usuellement recommandé de choisir un rayon de coupure $r_{\rm cut}$ supérieur ou égal à la distance aux premiers voisins, mais si possible excluant les seconds voisins.

On connaît donc l'expression finale de l'énergie DFTB si on arrête le développement au second ordre :

. .

$$E = \sum_{i=1}^{N_e} \sum_{AB} \sum_{\mu \in A} \sum_{\nu \in B} f_i c_{i\mu} c_{i\nu} H^0_{\mu\nu} + \frac{1}{2} \sum_{AB} \gamma_{AB} \left(\tau_A, \tau_B, R_{AB} \right) \Delta q_A \Delta q_B + E_{\text{rep}}.$$
 (3.16)

On en détermine le minimum par variation de $\delta (E - \sum_i \varepsilon_i \langle \psi_i | \psi_i \rangle)$, où le mutiplicateur de Lagrange ε_i est à déterminer. On note $S_{\mu\nu} = \langle \psi_\mu | \psi_\nu \rangle$ le recouvrement des orbitales μ et ν ; on veut donc, pour tout i et μ , résoudre un équivalent des équations de Kohn-Sham pour DFTB:

$$\sum_{\nu} c_{i\nu} \left(H_{\mu\nu} - \varepsilon_i S_{\mu\nu} \right) = 0, \qquad (3.17)$$

où $H_{\mu\nu} = H^0_{\mu\nu} + \frac{1}{2}S_{\mu\nu}\sum_C (\gamma_{AC} + \gamma_{BC})\Delta q_C$, A et B étant les atomes hôtes respectifs des orbitales μ et ν . En notant $v^e_A = \sum_B \gamma_{AB}\Delta q_B$ le potentiel électrostatique appliqué sur l'atome A, on peut réécrire

$$H_{\mu\nu} = H^0_{\mu\nu} + h^1_{\mu\nu} S_{\mu\nu} \text{ où } h^1_{\mu\nu} = \frac{v^e_A + v^e_B}{2} , \ \mu \in A , \ \nu \in B.$$

Similairement à leur analogue DFT, ces équations doivent être résolues itérativement, par calculs auto-cohérents (SCC). Comme nous l'avons vu plus haut, ce processus est inhérent aux termes en Δq , qui apparaissent lors du développement au second ordre; c'est pourquoi on réfère, parfois, aux calculs DFTB2 par l'appellation SCC-DFTB.

Enfin, intéressons nous au calcul des forces : pour cela, on calcule simplement le gradient de l'énergie totale (3.16) (et donc des trois termes constitutifs) par rapport aux

positions atomiques.

$$\vec{F}_A = -\sum_{i=1}^{N_e} \sum_B \sum_{\mu \in A} \sum_{\nu \in B} f_i c_{i,\mu} c_{i\nu} \left[\frac{\partial H^0_{\mu\nu}}{\partial \vec{r}_A} - \left(\varepsilon_i - h^1_{\mu\nu} \right) \frac{\partial S_{\mu\nu}}{\partial \vec{r}_A} \right] - \Delta q_A \sum_B \frac{\partial \gamma_{AB}}{\partial \vec{r}_A} \Delta q_B - \frac{\partial E_{\text{rep}}}{\partial \vec{r}_A}.$$
(3.18)

Chapitre 4

Objectif

En l'état, les potentiels répulsifs manipulés au sein de la méthode DFTB sont générés ainsi : il est nécessaire d'effectuer la différence entre des calculs DFT et DFTB sans interaction répulsive pour un jeu de systèmes et d'identifier l'impact de la distance entre deux atomes sur l'énergie de leur liaison. Il en résulte un potentiel de paire à portée limitée que l'on sait très bien utiliser en simulation atomistique. Le processus pour l'obtenir doit être effectué manuellement et donne un outil optimisé pour représenter au mieux le comportement d'une liaison chimique ; il vise à reproduire les énergies et structures des phases employées comme références. La variété de ces références est assez réduite. D'une part, on ne peut représenter qu'un type de liaison chimique par plage de distance ; exit donc les systèmes différents mais aux distances de liaisons analogues.



FIGURE 4.1 – Exemple de détermination de l'énergie répulsive d'une liaison C-H dans la molécule de méthane.

On peut par exemple penser au processus présenté précédemment dans la figure 3.2 du chapitre précédent, où l'on construit un potentiel répulsif de type *splines* pour la liaison C-C. Dans ce cas précis, on considère trois plages de distances : une très courte ou la liaison entre les atomes est triple ($r_{C\equiv C} < 1.25$ Å), une intermédiaire où la liaison est double ($r_{C=C} \in [1.25, 1.40]$ Å) et une plus longue ($r_{C-C} > 1.40$ Å) où la liaison est simple. Pour chacune de ces plages de distances, un unique type de liaison peut être décrit.

Pour la génération de potentiels répulsifs de paire à partir de systèmes isolés, il est parfois ardu de s'assurer que l'énergie répulsive calculée soit bien celle que l'on veut reproduire. Prenons l'exemple donné dans la figure 4.1, où l'on veut déterminer l'énergie répulsive de la liaison C-H dans la molécule de méthane pour $r \in [r_0, r_f]$. On choisit pour cela un atome (nommé ¹H dans la figure) : il s'agit du seul dont on modifie la position durant la procédure, les autres atomes sont figés. On veut

placer ¹H à $r = r_0$ de l'atome de carbone, puis augmenter graduellement r jusqu'à r_f .

Néanmoins, il est nécessaire de garantir que seule l'énergie répulsive entre ¹H et l'atome de carbone sera impactée par le déplacement. En d'autre termes, il faut s'assurer pour toutes les distances que l'interaction répulsive entre ¹H et tous les autres atomes d'hydrogène ⁿH, $n \ge 2$ soit rigoureusement nulle. En pratique, le choix d'un r_0 tel que cette condition se vérifie n'est pas aisé, notamment si la structure de la molécule de référence est complexe.

Enfin, certain écueils sont également à éviter lorsque l'on souhaite reproduire l'énergie répulsive d'une liaison au sein d'un cristal. Ainsi, lorsque les distances aux premiers et deuxièmes voisins sont assez proches, il est parfois ardu de déterminer quels atomes sont liés ou non par une interaction répulsive. C'est typiquement le cas pour la symétrie *cubique* face centrée, dont on présente l'aspect général dans la figure 4.2.

Usuellement, on détermine l'énergie répulsive par liaison des cristaux en mettant la boite de simulation à l'échelle d'un facteur variable, de manière à modifier d'autant les distances aux premiers voisins. Comme les distances aux premiers voisins sont identiques (r_1) , on espère ainsi évaluer l'énergie répulsive associée aux liaisons entre 1Au et nAu , $n \in [2, 4]$. Il est toutefois possible pour les systèmes très compressés que les seconds voisins (nAu , $n \in [5, 7]$, distants de r_2 dans la figure) deviennent suffisamment proches pour qu'une interaction répulsive émerge avec 1Au . Dès lors, on ne reproduit pas seulement l'énergie répulsive de la liaison, mais cette énergie parasitée par l'interaction avec les seconds voisins.

Enfin, le choix du système de référence (tel cristal ou telle molécule) pour une paire d'éléments chimiques donnée n'est pas unique; une large variété



FIGURE 4.2 – Structure cubique face centrée et distance aux premiers et seconds voisins pour l'or.

de candidats est en général disponible. Il est nécessaire de n'en conserver que quelques uns pour générer un potentiel de paire. Bien que l'on espère que ce choix soit peu important, c'est-à-dire que la transférabilité d'un système à l'autre soit garantie, ce n'est pas toujours exact.

Ainsi, la contrainte de reproduire l'énergie répulsive à l'échelle des liaisons est délicate : il est possible pour un utilisateur expérimenté de contourner ces écueils, mais ils posent une difficulté certaine. Même s'il est en théorie possible d'optimiser un potentiel de paire pour n'importe quel cas possible, il est trop ardu d'identifier l'énergie d'une liaison pour bon nombre de configurations : on doit en pratique se concentrer sur des systèmes nonpériodiques (pour lesquels on peut espérer faire varier la longueur d'une liaison seule) ou très symétriques (pour lesquels toutes les liaisons sont équivalentes) avec des seconds voisins assez éloignés. La capacité à représenter d'autres systèmes n'est pas acquise, et on peut aisément voir (Sec. 3.3, Fig. 3.3) que l'assomption d'une somme de contributions de paires s'avère parfois limitante. En s'affranchissant du concept d'énergie répulsive de liaison atomique, nous sommes en mesure d'outrepasser ce problème et d'employer n'importe quel système comme donnée de référence.

Pour cela, nous proposons d'utiliser des méthodes d'apprentissage automatique : leurs performances en interpolation dans de nombreux domaines - dont la simulation atomistique - sont notoires.

On dénombre d'ores et déjà des tentatives similaires appliquées à améliorer *DFTB* dans la littérature. Des propositions très proches de nos choix s'ébauchent parfois à demimot, comme des perspectives futures. Dans le cas de Ammothum Kandy *et al.* [37], le cœur du travail se concentre sur l'obtention de potentiels répulsifs de paire les plus transférables possible. L'idée est de trouver les paramètres optimaux (nombre d'intervalles employés pour les splines, paramètres de chacun d'entre eux) en imposant des contraintes formelles variées (signe ou monotonie par exemple). Si cette approche permet de rendre compte de nombreux polymorphes cristallins de manière individuelle, les auteurs concluent néanmoins que leur forme de paire empêche les potentiels répulsifs de couvrir simultanément une large variété de systèmes ; une approche avec un potentiel répulsif à plusieurs corps (très proche de celle que l'on adopte, présentée dans les chapitres suivants) est préconisée, motivée par quelques résultats prometteurs quant à la capacité de généralisation de tels outils.

À l'inverse, certains travaux très aboutis existent, et présentent des résultats variés. Par exemple, Van den Bossche *et al.* [38] proposent une procédure pour optimiser efficacement plusieurs potentiels répulsifs de paire spécifiquement sur **une** phase très précise d'un système donnée. La procédure s'inscrit dans la recherche de minima globaux de plusieurs cristaux variés (carbone, dioxyde de titane, oxyde de palladium, hydroxyde de calcium) et consiste pour chacun d'eux à optimiser conjointement la géométrie desdits cristaux et les paramètres des potentiels répulsifs de manière itérative. On arrive alors à des configurations très proches des minima effectifs DFT, avec des potentiels répulsifs de paire extrêmement performants pour ces configurations particulières. Cette approche par optimisation globale manipule des outils permettant de parfaire très efficacement des potentiels répulsifs de paire spécifiquement pour des systèmes donnés; ceux-ci sont construits pour être précis individuellement, et leur capacité de généralisation n'est ni investiguée ni réellement visée.

En parallèle, d'autres se concentrent uniquement sur le hamiltonien électronique, comme par exemple Haichen *et al.* [39]. L'idée est ici d'utiliser DFTB comme un intermédiaire pour prédire les propriétés électroniques de molécules. Dans ce cas, on conserve les interactions répulsives sous la forme d'un potentiel de type splines à deux corps, et l'on investigue deux modèles pour générer les éléments de matrice du hamiltonien : des termes de paire dépendant uniquement des espèces chimiques et des distances, et des réseaux de neurones tenant simultanément compte de tout l'environnement moléculaire à plusieurs corps. Les auteurs observent peu de différences entre les deux modèles, semblant indiquer d'après eux une limitation intrinsèque dans la forme même des hamiltoniens électroniques. Ils concluent notamment sur la possibilité d'inclure au sein du terme apparaissant au second ordre du développement DFTB la prise en compte de multipôles au-delà de l'ordre 2. On retient de cet article que l'emploi d'outils d'apprentissage automatique pour améliorer directement les paramétrisations des termes électroniques semble infructueux, et le potentiel répulsif paraît effectivement être un meilleur candidat à l'optimisation. Cette dernière intuition semble partiellement confirmée par Goldman *et al.* [40], dont les travaux s'intéressent à l'hydrogénation du plutonium. Ils s'appliquent à redéfinir le potentiel répulsif comme une combinaison de polynômes de Tchebychev [41] à **deux et trois corps**, l'ordre maximal des polynômes étant en principe aussi élevé que nécessaire. Les propriétés présentées semblent intéressantes, et des travaux plus récents [42] réinvestissent cette approche en la couplant à une optimisation conjointe des paramètres du hamiltonien électronique pour TiH₂. Les résultats semblent aussi spécifiques que prometteurs, mais sont malheureusement peu mis en perspective (par rapport aux potentiels de paire ou à d'autres approches). Les performances présentées justifient indéniablement la nécessité d'investigations plus poussées à l'avenir.

Au passage, il est intéressant de noter que les interactions à courte portée du formalisme ne sont pas uniquement régies par le potentiel répulsif. Ainsi, Wengert *et al.* [43] s'intéressent dans un papier récent à l'étude de cristaux moléculaires en apportant des corrections énergétiques aux courtes portées à DFTB, sans directement chercher à corriger l'énergie répulsive. Pour cela, le choix est fait d'utiliser des méthodes à noyaux (voir 5.3.2) pour traiter séparément des correction apportée aux niveaux intra- et inter-moléculaires. Ce choix s'avère pertinent pour les cristaux moléculaires, puisqu'il permet notamment pour ces derniers une bonne prédiction de leur stabilité relative.

Enfin, Kranz et al. [44] focalisent leurs travaux sur la génération de potentiels répulsifs à plusieurs corps pour une large gamme de systèmes moléculaires. Ils génèrent plusieurs potentiels répulsifs utilisant des outils d'apprentissage automatique présentés plus loin : la description des géométries est assurées par des matrices de Coulomb (voir 5.4.2.2) couplées à des méthodes à noyaux. Leur description s'articule autour du concept d'environnement de liaison atomique : on a donc un potentiel à plusieurs corps mais centré sur la liaison entre deux atomes, ce qui est pertinent dans le cas de molécules. L'emploi de tel ou tel potentiel particulier est déterminé via une approche par *clustering*, au sein de laquelle on détermine l'outil le plus adapté à la molécule étudiée en comparant sa géométrie à celles de référence. La transférabilité de cette approche est soit très bonne soit très mauvaise par définition, selon le sens exact que l'on accorde au terme *«transférabilité»*. Chaque potentiel est très spécifique et ne montre aucune preuve de généralisation, mais la méthode s'adapte aisément à de nouvelles liaisons via l'ajout de nouveaux potentiels répulsifs adaptés. La compatibilité avec des systèmes périodiques n'est pas détaillée, et le recours à plusieurs sous-potentiels en fonction de la situation rencontrée soulève également la question des cas limites, où le changement de potentiel peut induire d'éventuelles discontinuités. Le traitement des interactions répulsives au sein de molécules peut également se faire au travers d'autres méthodologies : on pense par exemple aux travaux récents de Panosetti et al. [45] utilisant la méthode Gaussian Process Regression (basée sur des méthodes à noyaux) ou à Stöhr et al. [46] recourant pour leur part à des réseaux de neurones profonds tensoriels (Deep Tensor Neural Network). Dans chacun de ces deux cas, les précisions accessibles pour les systèmes moléculaires sont notables mais semblent difficiles à adapter efficacement à des matériaux périodiques.

Notre approche cherche à se démarquer de toutes celles citées précédemment en insistant sur le point suivant : on désire un **unique** potentiel pour une paire donnée d'espèces chimiques. En pratique, on espère éviter les discontinuités induites par l'utilisation conjointe et conditionnelle de plusieurs potentiels. Nous souhaitons par conséquent couvrir tous les systèmes accessibles pour ces éléments chimiques, fussent-ils périodiques ou non, désordonnés ou non, stables ou hors équilibre. De plus, le potentiel répulsif enrobe une complexité théorique très forte au cœur du formalisme *DFTB*, mais sa forme d'interaction de paires repose plus sur une commodité pratique que sur une justification physique forte. Nous avons d'ailleurs vu que cette construction montre des limitations : il nous semble capital de profiter de la redéfinition mise en place dans les présents travaux pour la dépasser via une définition intrinsèque à plusieurs corps (c'est-à-dire de ne pas se limiter à un nombre fini de corps). Nous comptons explorer pour cette forme plus générale sa capacité à rendre compte de la physique complexe sous-jacente.

Le choix de ne reproduire avec les outils d'apprentissage automatique que l'énergie répulsive, et non l'énergie totale, présente plusieurs avantages majeurs : le plus notable est évidemment la capacité à conserver un accès aux propriétés électroniques du système. Par ailleurs, cette approche nous autorise à ignorer les interactions à longue portée, déléguées au reste du formalisme *DFTB*. Celui-ci permet de conserver une intuition physique forte, et joue grâce à sa robustesse un rôle de garde-fou dont une approche complètement paramétrique (c'est-à-dire ne reposant **que** sur des outils d'apprentissage automatique) ne bénéficie pas.

Notons au passage que ces dernières peinent de manière générale à être performantes simultanément sur les systèmes périodiques et non périodiques. Pour les systèmes périodiques, le traitement physique se fait usuellement à l'échelle des atomes (combien de voisins, nature des angles et des distances en jeu), parfois assistées de corrections aux longues distances. Pour les objets isolés, les états quantiques sont fortement localisés, à l'inverse des solides. La localisation ou non des états peut s'accompagner de changements de comportements notables. Par exemple, dans le cas d'états localisés, on peut plus facilement voir apparaître des charges ponctuelles, entraînant des interactions coulombiennes à longue portée, alors que des états délocalisés peuvent uniformiser l'état de charge du système, avec donc des interactions locales. En raison des différences induites par la (dé)localisation des états quantiques, la description des géométries est donc orientée en général vers l'une ou l'autre de ces classes de systèmes. Cette limitation est notamment discutée plus en détail dans [47], où les quelques lignes précédentes sont reformulées ainsi :

Current ML potentials deal with inorganic materials or molecules — but not normally with both $[\cdots]$ while there are connections between the two worlds, for example in the description of bulk liquid water, a truly unified ML potential framework that would at the same time treat a wide range of inorganic solids and organic molecules is not currently available. One of the challenges here is linked to the localized (delocalized) nature of electrons in molecules (periodic solids), respectively, and how this manifests in the different physical models and simplifications that can be made for them.

Des approches récentes prometteuses aspirent à tenir compte de cette difficulté en

Discussion des perspectives pour le domaine de la simulation atomistique par V.L. Deringer, M.A. Caro et G. Csànyi [47].

incluant dans leur représentation une équilibration globale des charges du système utilisée pour ajuster l'énergie calculée localement [48,49].

On espère pour notre part pouvoir dépasser cette problématique en laissant aux termes électroniques de DFTB la responsabilité des interactions longue distance et de l'état de charge, et ainsi pouvoir représenter convenablement ces deux classes de système avec le même outil.

Dans l'idéal, le changement apporté par les présents travaux vise évidemment à améliorer la précision de *DFTB* vis-à-vis de la théorie de la fonctionnelle de la densité, pour une plus large gamme de systèmes que jusqu'alors. La quantification exacte de cette précision est dépendante des systèmes considérés : géométries d'objets isolés, stabilité relative pour les cristaux, distributions radiales et angulaires, propriétés vibrationnelles...

Il est également souhaitable de faciliter la génération d'interactions répulsives, en la basant sur un processus défini reposant sur décisions clé de l'utilisateur, justifiées et identifiées.

En parallèle, une forte emphase est mise sur la performance numérique : le faible coût de calcul des méthodes DFTB en comparaison des méthodes premier principe est d'ailleurs leur principal intérêt. Le temps de calcul du potentiel répulsif de paire est négligeable par rapport à celui de la partie électronique, et l'on souhaite que l'alternative présentée conserve cette caractéristique.

Chapitre 5

Apprentissage automatique

Si le concept de *machine-learning* est aujourd'hui profondément implanté dans le quotidien de tous, en donner une définition spécifique semble être une entreprise ardue. Cette science, dont la francisation en **apprentissage automatique** s'impose peu à peu dans l'usage, regroupe en réalité une zoologie aux limites floues de méthodes très variées. Intrinsèquement, elles ont cependant quelques points communs : elles visent à réaliser des tâches pour lesquelles elles ne sont pas explicitement programmées, se fondent sur des modèles statistiques basés sur l'inférence et reposent donc sur l'ingestion puis l'assimilation de données.

On trouve d'ailleurs en ce dernier point une piste d'explication quant à leur omniprésence actuelle. Bien qu'étant historiquement aussi vieilles que l'ordinateur (et reposant sur des fondations qui lui sont même antérieures), l'explosion des données disponibles causée par le progrès conjoint des technologies informatiques et d'Internet a désigné *de facto* les méthodes d'apprentissage automatique comme des candidates de choix pour traiter quantité de problèmes, à toutes les échelles de la société.

Dans cette partie, qui ne se veut évidemment pas exhaustive, nous présentons le concept d'apprentissage automatique, son histoire ainsi que les méthodes pertinentes dans le cadre de cette thèse.

5.1 Présentation historique

Dès le $18^{\grave{e}me}$ siècle, les travaux de Thomas Bayes [50] posent les premières fondations formelles de l'apprentissage automatique, lesquelles constituent la base du théorème de Bayes, présenté pour la première fois par Pierre-Simon Laplace début $19^{\grave{e}me}$ siècle [51]. Ce théorème décrit la probabilité d'un évènement en se basant sur nos connaissances *a priori*.

La formalisation de l'idée de connaissance *a priori* est d'ailleurs une étape importante dans le développement de la théorie de la probabilité. Cela donne naissance au concept d'inférence bayésienne : il s'agit de l'adaptation des *a priori* en fonction des preuves et informations rencontrées. Cette approche, très similaire au fonctionnement du raisonnement humain, forme le cœur des méthodes d'apprentissage automatique et explique pourquoi elles sont parfois confondues avec les *intelligences artificielles*.

Les travaux d'Alan Turing constituent évidemment la clé de voûte de l'apprentissage automatique tel que nous le connaissons aujourd'hui. Au-delà de la paternité de l'ancêtre de l'ordinateur avec sa *machine de Turing* durant la seconde guerre mondiale, il propose également une *machine apprenante* ayant la capacité de devenir artificiellement intelligente [52].

Contemporainement à la machine de Turing, les premiers réseaux de neurones font leur apparition [53], nommés ainsi pour leur fonctionnement en principe similaire aux neurones biologiques. Cette analogie anatomique leur vaut d'ailleurs une place de choix dans l'imaginaire collectif; l'intérêt qu'ils cultivent, couplé aux possibilités offertes par leurs qualités d'approximateurs universels [54] permettent aux modèles de s'enrichir grandement. On cite par exemple l'apparition du *perceptron* [55], des réseaux de neurones à convolution [56], de l'algorithme de rétropropagation [57], des réseaux de neurones récurrents [58], *etc.* On se propose de présenter plus en détail les réseaux de neurones plus loin.

En parallèle des avancées connues par les réseaux de neurones, d'autres outils d'apprentissage automatique voient le jour. On pense par exemple à l'apprentissage par renforcement [59], aux forêts d'arbres décisionnels [60], aux machines à vecteurs de support [61] ou encore aux méthodes à noyaux [62].

Enfin, on peut citer quelques accomplissements, jouant aux yeux de l'opinion publique le rôle de pierres angulaires dans la progression des méthodes d'apprentissage automatique. Il s'agit, en général, de performances de machines comparables ou supérieures à des humains pour des tâches spécifiques. Les plus notables se trouvent dans le domaine du jeu : la capacité de jouer aux dames (1952, [63]), au morpion (1963, [64]) ou même de battre des champions humains aux échecs (1997, [65]) ou plus récemment au jeu de Go (2016, [66]). Si ces étapes ont une portée plus symbolique que pragmatique, on peut également noter l'apprentissage de la prononciation (1985, [67]), la détection d'animaux (2012, [68]) ou de visages (2014, [69]) en direct sur support vidéo.

5.2 Applications en simulation atomistique

Parmi la pléthore d'applications où les outils d'apprentissage automatique trouvent une utilité, on s'intéresse ici spécifiquement au domaine de la simulation atomistique.

En étudiant la matière à l'échelle d'atomes, on espère pouvoir représenter des comportements fins avec un contrôle extensif sur le système ; la qualité de ce genre d'informations, difficilement accessibles expérimentalement, est néanmoins très dépendante de la méthode de calcul utilisée (Fig. 3.1).

Dans chacune de ces méthodes, il existe de nombreux candidats pour lesquels la flexibilité et l'efficacité des méthodes d'apprentissage automatique au sens large offrent

des perspectives prometteuses.

Dans cette section, on se propose de présenter, de manière succincte et non exhaustive, quelques applications pour lesquelles l'intégration d'outils d'apprentissage automatique en simulation atomistique permet un gain de précision et/ou de performance notable.

Comme point de départ, il semble pertinent d'aborder la principale jonction entre ces deux sciences : les *potentiels numériques*. Afin de mieux percevoir les enjeux de cette approche, nous avons besoin d'un peu de contexte.

D'une part, nous avons présenté précédemment (Chap. 2) les intérêts, les performances et les limitations des méthodes premier principe : en se basant sur les équations de la mécanique quantique (2.1) et en utilisant un minimum d'approximations, on espère pouvoir évaluer très précisément les propriétés du système considéré, au prix d'un coût de calcul élevé.

Par ailleurs, on peut à l'inverse tenter de simplifier fortement la représentation des interactions du système afin de diminuer drastiquement le coût des calculs : on s'affranchit des propriétés électroniques et on représente les interactions entre atomes par des fonctions de forme définie, avec quelques paramètres et dépendant uniquement des positions atomiques. On parle alors de potentiels empiriques, dont la figure de proue est le potentiel de *Lennard-Jones*, proposé par *John Lennard-Jones* en 1931 [70] et permettant de représenter efficacement le comportement de fluides.

$$E(r) = -4\varepsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right].$$

S'ils reposent sur des approximations pertinentes, visant à rendre compte de phénomènes physiques bien connus, leurs paramètres sont au bout du compte ajustés afin de reproduire des données de référence, expérimentales ou issues de calculs *premier principe*.

Ces potentiels sous forme de fonctions paramétrées sont à double tranchant. Leur forme fonctionnelle prédéfinie permet d'avoir un nombre réduit de paramètres ayant une signification physique et assure, *a minima*, un comportement physique cohérent. En revanche, elles présentent intrinsèquement une transférabilité réduite : ces paramètres sont ajustés pour une gamme spécifique de systèmes, et en reproduisent convenablement un nombre fini de propriétés.

C'est exactement dans cette optique que l'intégration d'outils d'apprentissage automatique est pertinente : si la forme prédéfinie des potentiels empiriques et leur nombre réduit de paramètres ajustés à la main sont limitants, les potentiels numériques proposent de s'en affranchir. Ils emploient pour cela les méthodes d'apprentissage automatique, à la fondation physique moindre mais très flexibles car manipulant beaucoup plus de paramètres, et relativement peu coûteuses numériquement. Ce choix nécessite évidemment une attention toute particulière pour s'assurer que les comportements produits ont un sens physique en renforçant le formalisme des outils d'apprentissage automatique par quelques intuitions physiques (voir 5.4). Néanmoins, cette méthodologie permet en principe d'obtenir des potentiels s'approchant de la précision des méthodes premier principe et applicables aux mêmes échelles que des potentiels empiriques (Fig. 5.1).



Transférables

FIGURE 5.1 – Présentation schématique du concept de potentiel numérique.

La principale limitation intrinsèque des potentiels purement numériques réside dans leur incapacité à fournir des informations sur l'état électronique du système [71]. Si l'on se place du point de vue du coût numérique, la désintrication du calcul de structure électronique coûteux et de la précision des méthodes premier principe est un avantage : les utilisateurs réguliers de potentiels empiriques trouvent un intérêt pour cette approche. En revanche, les calculs de structure électronique rendent disponibles nombre de propriétés inaccessibles autrement.

Par ailleurs, on emploie parfois des outils d'apprentissage automatique pour améliorer les résultats obtenus avec des méthodes premier principe. Si cela peut paraître curieux de prime abord, il faut commencer par rappeler que (Chap. 2) toutes les méthodes que l'on nomme *premier principe* par abus de langage comportent des approximations, bien que raisonnables et en nombre restreint, qui leur sont propres. Toute approximation étant intrinsèquement plus performante dans des conditions spécifiques, on peut trouver pour chaque méthode des situations où il y a place à amélioration.

Si l'on ne se propose évidemment pas d'en faire une liste exhaustive, faute de place et de compétence, donnons en ici un exemple pertinent : l'amélioration de la capacité de la théorie de la fonctionnelle de la densité (DFT, voir 2.3) à représenter les systèmes moléculaires, ici l'eau. Le détail de cette approche peut être trouvé dans la littérature [72].

Les systèmes moléculaires sont lourdement constitués d'interactions d'échange et corrélation et d'interactions dépendantes de la distribution de charges dans le système, autant de comportements représentant une réelle difficulté pour la méthode *DFT*. Dès lors, l'idée est de s'appuyer sur des données obtenues via une autre méthode premier principe de chimie quantique corrélée (CCSD(T), [73]), bien mieux adaptée à ce genre de systèmes mais plus coûteuse et moins utilisée. On cherche ensuite à entraîner des outils d'apprentissage automatique (basés sur des méthodes à noyaux - voir 5.3.2) à corriger les fonctionnelles hybrides (2.3.1, 2.3.2, ...) de DFT pour reproduire le comportement réel des systèmes moléculaires. Dans ce cas précis, la rectification des fonctionnelles s'effectue via la correction des erreurs commises par leurs termes à un et deux corps; ceci apporte un gain notable de précision, plafonné néanmoins par la non considération des termes audelà de deux corps. In fine, l'idée reste d'obtenir une paramétrisation DFT performante qui demeure, quoi qu'il en soit, relativement coûteuse numériquement. La restriction à la correction d'un terme spécifique plutôt que la prise en charge de l'intégralité des énergies par les outils d'apprentissage automatique demeure néanmoins pertinente : on conserve d'une part accès aux informations électroniques, et d'autre part la robustesse certaine du formalisme DFT permet de répliquer cette approche aisément à de nouveaux systèmes.

5.3 Outils d'apprentissage

Si les quelques sections précédentes permettent d'avoir une vue d'ensemble raisonnable de l'apprentissage automatique, d'un point de vue général ou plus spécifiquement couplé à la simulation atomistique, il semble intéressant de présenter plus en détail les fonctionnements précis et le formalismes associés à quelques uns de ces outils.

Nous détaillons donc le fonctionnement des réseaux de neurones à propagation avant, utilisés dans le cadre de cette thèse, mais également un outil pertinent envisagé en tant qu'alternative : les méthodes à noyaux.

5.3.1 Réseaux de neurones

Un neurone artificiel est une fonction mathématique recevant une ou plusieurs entrées et appliquant une transformation pour retourner une valeur en sortie. Ce fonctionnement est très similaire à celui des neurones biologiques constitutifs des cerveaux animaux, dont les nôtres.

La nature de la transformation appliquée est double. D'une part, le neurone attribue un certain poids à chacune des N entrées x: on appelle les paramètres du neurone *poids* w et *biais b*. Cette première manipulation permet d'obtenir l'*activation a* du neurone.

$$a = b + \sum_{i=1}^{N} w_i x_i.$$
 (5.1)

D'autre part, le neurone transmet son activation a à sa fonction d'activation (ou fonction de transfert) pour obtenir sa valeur de sortie y.

$$y = f(a). \tag{5.2}$$

Cette fonction peut varier en fonction de l'application visée : si l'objectif est une classification binaire on peut choisir une fonction seuil discontinue, alors que pour d'autres applications une fonction continue de forme variable est préférable. Voici quelques exemples de fonctions de transfert :

$$f(x) = x$$
 (5.3a) $f(x) = \tanh(x)$ (5.3b)

$$f(x) = \frac{1}{1 + e^{-x}} \qquad (5.3c) \qquad \qquad f(x) = \begin{cases} 0 & , \ x < 0 \\ 1 & , \ x \ge 0 \end{cases} \tag{5.3d}$$

Cette brique élémentaire est représentée schématiquement dans la figure 5.2.



FIGURE 5.2 – Schéma du fonctionnement d'un neurone artificiel.

Dès lors, la définition d'un réseau de neurones est triviale : il s'agit d'une combinaison de neurones communiquant entre eux, la sortie de certains étant l'une des entrées d'autres. On peut alors imaginer deux classes de réseaux de neurones :

- les réseaux où les valeurs d'entrée des neurones sont indépendantes de leur valeur de sortie; c'est-à-dire les réseaux sans rétroaction où l'information se propage dans une seule direction. On parle alors de **réseaux de neurones à propagation avant** (*feed-forward neural networks*), comme par exemple dans la figure 5.3a.
- les réseaux où une dépendance cyclique apparaît; c'est-à-dire les réseaux où les valeurs d'entrée d'un ou plusieurs neurones peuvent dépendre de leur sortie. On parle alors de **réseaux de neurones récurrents** (*recurrent neural network*), comme par exemple dans la figure 5.3b.



(a) Exemple de réseau de neurones à propagation avant.



(b) Exemple de réseau de neurones récurrent.

FIGURE 5.3 – Différence entre les deux classes de réseaux de neurones.

Le choix du type de réseau est dépendant de l'application visée. Un réseau de neurones à propagation avant, à partir de valeurs d'entrée, fournit des valeurs de sortie : on peut le visualiser comme une application F^p qui à un vecteur \vec{X} associe un vecteur \vec{Y} , indépendamment des vecteurs qu'on lui a donné précédemment.

$$F^p: \vec{X} \mapsto \vec{Y} = F^p(\vec{X}).$$

A l'inverse, la structure d'un réseau de neurones récurrent lui confère une forme de *mémoire interne* lui permettant de traiter des séquences d'entrées de taille prédéfinie N. Il s'agit donc d'une application F^r qui retourne un vecteur de sortie \vec{Y} en fonction du vecteur présenté \vec{X}_i et des $\{\vec{X}_s\}_{s=i-N}^{i-1}$ vecteurs d'entrée précédents :

$$F^{r}: \vec{X}_{i}, \left\{\vec{X}_{s}\right\}_{s=i-N}^{i-1} \mapsto \vec{Y} = F^{r}\left(\vec{X}_{i-N}, \cdots, \vec{X}_{i-1}, \vec{X}_{i}\right).$$

Autrement dit, un réseau de neurones récurrent n'associe pas toujours la même sortie à la même entrée. Ce genre de comportement est extrêmement utile dans de nombreuses applications de la vie courante comme la reconnaissance vocale, ou celle de textes écrits à la main ou de vidéos. L'aide à la saisie de texte constitue un exemple visuel et pertinent : en fonction de la séquence de lettres et de mots saisie par l'utilisateur, l'outil suggère une suite adaptée au contexte. On peut facilement en faire l'expérience avec un téléphone (Fig. 5.4), où la saisie d'un même mot n'entraîne pas systématiquement une unique suggestion, mais aspire à s'adapter au contexte en fonction des quelques autres employés directement avant.



FIGURE 5.4 – Exemple d'utilisation courante de réseaux de neurones récurrents.

Pour les travaux effectués dans le cadre de cette thèse, ce genre de *mémoire* n'est pas souhaitable et l'on choisit d'utiliser uniquement des réseaux de neurones à propagation avant. Dans la suite, seul ce type de réseau sera étudié et on utilisera l'appellation *réseau(x) de neurones* pour y faire référence, afin d'alléger les notations. De plus, l'abréviation NN (d'après l'anglicisme *neural network*) est parfois employée, notamment pour référer aux grandeurs liées aux réseaux de neurones dans diverses équations et figures.

Usuellement, on organise les réseaux par couches successives, les neurones de chaque couche prenant comme valeurs d'entrée les sorties des neurones de la couche précédente. On peut dès lors désigner :

• une couche d'entrée, dont les neurones ont chacun comme unique valeur d'entrée un élément X_i du vecteur d'entrée du réseau \vec{X} .

- une couche de sortie, dont les neurones retournent chacun en valeur de sortie un élément Y_j du vecteur de sortie du réseau \vec{Y} .
- des couches intermédiaires optionnelles, plus couramment appelées couches cachées, faisant le lien entre les couches d'entrées et de sortie. On leur réfère parfois en employant l'abréviation *hl* (d'après l'anglicisme *hidden layers*) dans certaines équations des pages à venir.

Dans de tels réseaux, la couche d'entrée a donc autant de neurones qu'il y a d'éléments dans le vecteur d'entrée \vec{X} , et la couche de sortie en a autant qu'il y a d'éléments dans le vecteur de sortie \vec{Y} .

En revanche, le nombre de couches cachées est arbitraire : on peut n'en avoir aucune, quelques unes (comme dans le cas de cette thèse), ou un très grand nombre - auquel cas on parle de réseaux de neurones profonds ou *deep learning*. Le nombre de neurones par couches cachées est également arbitraire, et toutes les couches cachées d'un même réseau n'ont pas nécessairement le même nombre de neurones.

On peut schématiser le fonctionnement de tels réseaux de neurones multicouches comme dans la figure 5.5, qui représente un réseau de forme 4-2-3-1 : c'est-à-dire un réseau $\mathbb{R}^4 \mapsto \mathbb{R}^1$ avec 4 neurones d'entrées, une première couche cachée contenant 2 neurones, une seconde couche cachée de 3 neurones et une couche de sortie d'un seul neurone.





Pour un réseau de neurones sans récurrence et organisé en couches de neurones successives, on peut donc exprimer le vecteur de sortie \vec{Y} en fonction du vecteur d'entrée \vec{X} comme :

$$\vec{Y} = f_{act}^{out} \left(\sum_{i=1}^{N_n^{hl}} w_{mi}^{out} \cdot f_{act}^{hl} \left(\sum_{j=1}^{N_n^{hl-1}} w_{ij}^{hl} \cdot \dots f_{act}^2 \left(\sum_{q=1}^{N_n^{in}} w_{pq}^2 \cdot X_q + b_p^2 \right) \dots + b_i^{hl} \right) + b_m^{out} \right).$$
(5.4)

En utilisant une formulation plus proche du fonctionnement en couches successives des réseaux de neurones, on arrive à une notation plus intuitive. Pour cela, on considère chaque couche indépendamment, et l'on note $\vec{Y^l}$ le vecteur de sortie de chaque couche l du réseau :

$$\begin{split} Y_{i}^{1} &= X_{i} \\ Y_{j}^{2} &= f_{act}^{2} \left(\sum_{i=1}^{N_{n}^{in}} w_{ji}^{2} \cdot Y_{i}^{1} + b_{j}^{2} \right) \\ Y_{k}^{3} &= f_{act}^{3} \left(\sum_{j=1}^{N_{n}^{2}} w_{kj}^{3} \cdot Y_{j}^{2} + b_{k}^{3} \right) \\ \cdots \\ Y_{q}^{l+1} &= f_{act}^{l+1} \left(\sum_{p=1}^{N_{n}^{l}} w_{qp}^{l+1} \cdot Y_{p}^{l} + b_{q}^{l+1} \right) \\ \cdots \\ Y_{m}^{out} &= f_{act}^{out} \left(\sum_{z=1}^{N_{n}^{hl}} w_{mz}^{out} Y_{z}^{hl} + b_{m}^{out} \right). \end{split}$$

On détermine le nombre de paramètres $N = N_w + N_b$ d'un réseau de neurones à N_l couches à partir de sa structure. Le nombre de poids N_w se calcule ainsi

$$N_w = \sum_{l=1}^{N_l} N_n^{l+1} \cdot N_n^l,$$

et le nombre de biais

$$N_b = \sum_{l=1}^{N_l} N_n^l.$$

Par exemple, le réseau de forme 4-2-3-1 présenté figure 5.5 a $N_w = 8 + 6 + 3 = 17$ poids et $N_b = 2 + 3 + 1 = 6$ biais. Ainsi, même un réseau très simple et relativement réduit présente tout de même un total de 23 paramètres.

Lorsque l'on parle d'*entraîner* un réseau de neurones (ou d'*apprentissage* pour un réseau de neurones), on désigne un processus où l'on cherche à déterminer les N valeurs de paramètres permettant de représenter au mieux un jeu de données de référence. Pour

nous, il s'agit de données $\{\vec{X}, \vec{Y}\}$ dont on sait quel vecteur \vec{Y} est associé à un vecteur d'entrée \vec{X} : on parle alors d'apprentissage supervisé. On veut donc les poids \vec{w} et biais \vec{b} tels que

$$\min_{\vec{w},\vec{b}} \left\{ \vec{Y}^{out} \left(\vec{X}, \vec{w}, \vec{b} \right) - \vec{Y} \right\}.$$
(5.5)

Le choix de la structure du réseau est en principe arbitraire, et on peut ajouter neurones et couches cachées à volonté, augmentant par la même occasion le nombre de paramètres et donc la flexibilité du réseau. Si cela peut paraître contre intuitif au premier abord, une trop grande flexibilité est cependant à proscrire. Ainsi, il est possible, même courant, qu'un réseau de neurones soit trop flexible, au point de reproduire trop fidèlement les données de référence jusqu'à leur bruit statistique, perdant ainsi de son pouvoir de généralisation : on parle alors de sur-apprentissage (Fig. 5.6).



FIGURE 5.6 – Schématisation du phénomène de sur-apprentissage.

Afin de détecter et prévenir ce phénomène, il existe plusieurs bonnes pratiques recommandées. Dans un premier temps, en amont de tout apprentissage, il est souhaitable de choisir la structure du réseau la plus petite possible, en limitant le nombre de neurones et de couches cachées au minimum nécessaire pour que le réseau parvienne à rendre compte des comportements étudiés. On prévient ainsi toute flexibilité superflue. Dans la pratique, ce choix de structure minimale n'est que très rarement trivial, et est souvent déterminé par un processus répété d'*essai-erreur* jusqu'à obtenir une structure convenable.

D'autre part, il est fréquent de répartir aléatoirement les données de référence disponibles en plusieurs sous-jeux. Un premier, usuellement nommé jeu d'apprentissage (*learning set*), est celui que l'on fournit au réseau pour effectuer l'apprentissage (c'est-à-dire l'optimisation des paramètres). Un second, que l'on nomme jeu de test (*testing set*), est utilisé sporadiquement **lors de l'apprentissage** comme outil de vérification : le réseau ne l'utilise pas pour optimiser ses paramètres, mais simplement pour comparer la précision des jeux d'apprentissage et de test, vérifiant par la même occasion que le réseau est capable de généraliser à des données qu'il ne *«connaît»* pas.



Dès lors que les précisions du réseau sur ces deux jeux de données divergent, et que l'erreur sur le jeu de test est significativement plus élevée que sur le jeu d'apprentissage, il y a sur-apprentissage (Fig. 5.7). Enfin, on peut optionnellement construire un troisième jeu de données, que l'on nomme jeu de validation (*validation set*), utilisé **après** optimisation, de la même manière que le jeu de test, pour renforcer un peu plus notre confiance dans la capacité du réseau à généraliser.

Bien que potentiellement utile, le jeu de validation a une fonction très similaire au jeu de test, et séparer en trois plutôt qu'en deux le jeu de données de références peut être contraignant si sa taille est

FIGURE 5.7 – Détecter le sur-apprentissage.

limitée. Dans le cadre de cette thèse, et des applications liant apprentissage automatique et simulation atomistique portées à notre connaissance, le choix est fait de travailler uniquement avec deux jeux de données : un jeu d'apprentissage et un jeu de test, dont les tailles relatives peuvent varier.

Évidemment, la construction du jeu d'apprentissage est capitale pour la précision du réseau : celui-ci doit être suffisamment complet pour être représentatif des systèmes que l'on souhaite couvrir, contenir peu de redondances et idéalement être de taille réduite, sans quoi le processus d'apprentissage peut être significativement long.

Profitons-en d'ailleurs pour discuter des facteurs influant le coût numérique des deux modes de fonctionnement classiques des réseaux de neurones, c'est-à-dire les processus de calcul (*compute*) où l'on détermine \vec{Y} connaissant \vec{X} , et d'apprentissage (*training*) où l'on optimise les paramètres pour reproduire les $\{\vec{Y}\}$ associés aux $\{\vec{X}\}$ du jeu de données d'apprentissage.

La forme du réseau, et donc le nombre de paramètres, mais aussi l'algorithme d'optimisation choisi (Sec. 6.2) et la taille du jeu de données d'apprentissage déterminent le coût numérique du processus d'apprentissage, et donc sa durée.

En revanche, lors du processus de calcul, seule la structure d'un réseau, c'est-à-dire le nombre de couches et de neurones qu'il contient, détermine le coût numérique. Si cette caractéristique peut sembler évidente dans le cas des réseaux de neurones, il existe certaines méthodes d'apprentissage automatique pour lesquelles ce n'est pas le cas, comme par exemple les méthodes à noyaux, présentées ci-après.

5.3.2 Méthodes à noyaux

Lors des travaux de cette thèse, une autre méthode d'apprentissage automatique est envisagée comme alternative aux réseaux de neurones : les méthodes à noyaux (*kernel methods*). La présentation de cet outil, plus superficielle, vise d'une part à en donner une explication suffisante pour comprendre les différents enjeux et raisonnements justifiant qu'on lui préfère des réseaux de neurones. D'autre part, on insiste sur quelques caractéristiques clés de cette approche, comme le temps de calcul, pour permettre de mieux interpréter le chapitre 8 des présents travaux, au sein duquel un potentiel adoptant cette approche [74] est utilisé comme point de comparaison.

D'un point de vue mathématique, un noyau est un objet permettant de mesurer la non-injectivité¹ d'un morphisme². Cet objet est intéressant puisqu'il établit une relation entre les éléments qui sont envoyés sur la même image par le morphisme. Grossièrement, on peut dire qu'un noyau permet une mesure de similarité entre deux objets issus d'une même application.

Afin d'avoir un aperçu rapide des méthodes à noyaux, il est intéressant d'introduire brièvement un élément central de la théorie de l'interpolation : les **RKHS** (*Reproducing Kernel Hilbert Spaces*). L'idée des RKHS est globalement de définir un espace d'interpolation très grand (de dimension potentiellement infinie) avec une structure sous-jacente permettant d'exprimer simplement l'interpolation. Si on considère un espace compact Xqui représente l'entrée, on peut définir un noyau $K : X \times X \mapsto \mathbb{R}$ vérifiant :

- une relation de symétrie : $\forall (x, x') \in X^2, K(x, x') = K(x', x).$
- que sa matrice de corrélation K est définie positive : $\forall N \in \mathbb{N}^*, (x_1, \cdots, x_N) \in X^N, \quad \forall (a_1, \cdots, a_N) \in \mathbb{R}^N, \quad \sum_{i=1}^N \sum_{j=1}^N a_i a_j K(x_i, x_j) \ge 0.$

Cette mesure de corrélation peut être interprétée comme un produit scalaire sur un espace de Hilbert, et on dit qu'il s'agit d'un *noyau reproduisant les espaces de Hilbert* (RKHS).

Cela implique que n'importe quelle mesure de similarité K définit un espace d'approximation : si l'on sait exprimer une distance entre deux objets d'entrée (similarité, corrélation, ...), on peut définir un espace d'approximation sur lequel repose la méthode à noyau.

Dans notre cas, on peut simplifier en écrivant que cet espace de dimension arbitraire est défini par une base d'apprentissage de la même dimension, et sur lequel on peut projeter grâce à la méthode à noyaux tout nouvel élément (symétrisé comme la base d'apprentissage) afin de connaître les similarités qu'il a avec les données de référence. Les ressemblances entre cet élément inconnu et ceux du jeu de données de référence sont utilisées pour déterminer le vecteur de sortie à associer.

Un exemple courant de noyau, notamment souvent utilisé en simulation atomistique,

^{1.} On dit qu'une application est une injection si tout élément de son ensemble d'arrivée a au plus un antécédent par cette application.

^{2.} Application qui relie deux ensembles (groupes, espaces vectoriels,...), par exemple une fonction.

est le noyau gaussien (5.6) [71].

$$K(x,y) = e^{-\frac{(x-y)^2}{2\sigma^2}}.$$
(5.6)

Un tel noyau permet de déterminer la sortie \vec{Y}_A à associer au vecteur d'entrée \vec{X}_A d'un système A inconnu. Pour cela, on peut utiliser la méthode kernel ridge regression, dont une application à la simulation atomistique est présentée dans [75].

Soit Ω le jeu contenant les *n* données de référence définissant l'espace sur lequel repose la méthode à noyaux, Φ l'application permettant d'associer un vecteur d'entrée \vec{X} au système considéré (voir 5.4). Le vecteur de sortie \vec{Y} calculé par la méthode à noyau gaussien est donné par l'équation (5.7),

$$\vec{Y}_{A} = \sum_{i=1}^{n} \alpha_{i}^{*} K\left(\Phi\left(\Omega_{i}\right), \Phi\left(A\right)\right) = \sum_{i=1}^{n} \alpha_{i}^{*} K\left(\vec{X}_{i}, \vec{X}_{A}\right), \qquad (5.7)$$

où $\vec{\alpha^*}$ est le vecteur de paramètres de la méthode à noyau, analogue aux poids et biais d'un réseau de neurones.

L'un des principaux avantages de cette méthode est la simplicité du processus d'optimisation, c'est-à-dire que la détermination de $\vec{\alpha^*}$ est simple et consistante : il est obtenu en résolvant le système matriciel (5.8),

$$\vec{\alpha^*} = (K + n\lambda I_d) \cdot M, \tag{5.8}$$

où $M = \left\{\vec{Y}_i\right\}_{i=1}^n$ est la matrice contenant les vecteurs de sortie du jeu de données de référence. K est une matrice (parfois appelée matrice de covariance) dont les éléments $K_{ij} = K(\vec{X}_i, \vec{X}_j)$ correspondent à l'application du noyau entre les vecteurs d'entrée \vec{X}_i et \vec{X}_j du jeu de données de référence. Enfin λ est un hyper-paramètre faisant office de coefficient de régularisation, fixé par l'utilisateur et pouvant être comparé à un bruit d'échantillonnage.

Remarque : de manière générale, toutes les méthodes à noyau nécessitent la construction d'une matrice de covariance sur la base d'apprentissage. La méthode de régression pour calculer les poids varie cependant (inversion de système matriciel, minimisation de fonction coût, ...).

Chaque jeu de données de référence induit un vecteur poids associé. Pour déterminer la valeur à associer à un nouvel élément inconnu, on a besoin du vecteur poids calculé $\vec{\alpha^*}$ et des vecteurs d'entrée $\{\vec{X}\}$ et de sortie $\{\vec{Y}\}$ (sous la forme de la matrice de covariance M) du jeu de données de référence; il n'est pas nécessaire de recalculer la matrice de covariance ou d'inverser le système matriciel (5.8) pour chaque nouvel élément.

Cette caractéristique seule constitue, dans le cadre cette thèse, une raison suffisante pour choisir de ne pas utiliser de méthodes à noyaux, car elle entraîne plusieurs conséquences notables. D'une part, cela implique que le coût numérique du processus de calcul (5.7) d'un vecteur de sortie \vec{Y} à partir d'un vecteur d'entrée \vec{X} dépend explicitement de la taille du jeu de données de référence : plus le jeu de données est grand, plus le calcul est long. Dans un code tel que DFTB+ où une forte attention est portée sur la performance, un tel comportement peut être préjudiciable.

D'autre part, on se heurte à un problème de portabilité. Dans DFTB+, les informations sur les interactions sont stockées et fournies au code dans des fichiers textes nommés fichiers de Slater-Koster (voir annexe C.1.1). Utiliser des méthodes à noyaux signifierait donc inclure dans ces fichiers l'intégralité de l'information sur le vecteur de paramètres α^* mais également sur les vecteurs d'entrée $\{\vec{X}\}$ et de la matrice de covariance M du jeu de données de référence. Une telle approche alourdit considérablement la quantité d'information contenue dans ces fichiers, amoindrissant la portabilité car compliquant le transfert entre utilisateurs.

5.4 Décrire l'environnement atomique

Nous avons justifié dans la section précédente (5.3) notre choix d'utiliser des réseaux de neurones. Ceux-ci offrent une capacité d'interpolation appréciable, mais comportent certaines limitations intrinsèques les rendant inaptes à traiter seuls n'importe quel problème. Le plus souvent, il est impossible de fournir les données brutes au réseau, et l'on doit les manipuler de sorte à respecter certaines contraintes. Cela requiert donc un prétraitement des données, que l'on appelle usuellement **extraction de caractéristiques** (*feature extraction*).

Certaines de ces contraintes peuvent être techniques. Prenons l'exemple de MNIST [76], où l'objectif est d'apprendre à reconnaître des chiffres manuscrits. Une image brute du chiffre ne peut pas être fournie en entrée du réseau de neurones : le vecteur d'entrée d'un réseau de neurones doit être consistant d'un système à l'autre, et donc de taille constante.

Il faut dans un premier temps normaliser l'image à une taille fixe (dans ce cas 28×28 pixels). On convertit ensuite cette image en un vecteur contenant une valeur numérique entre 0 et 1 dont chaque élément correspond à l'intensité d'un pixel, en niveau de gris (Fig. 5.8).



FIGURE 5.8 – Extraction de caractéristiques - exemple de MNIST.

Dans d'autres cas, tels que l'application qui nous intéresse, ces contraintes peuvent être conceptuelles. L'idée est de calculer l'énergie répulsive totale d'une boîte de simulation comme la somme des contributions de chaque atome.

$$E_{\rm tot}^{\rm rep} = \sum_{i=1}^{N_{\rm atomes}} E_i^{\rm rep}.$$
(5.9)

On veut donc, pour chaque atome, être capable grâce à un réseau de neurones de calculer E_i^{rep} . Nous avons vu précédemment (Sec. 3.1) que les interactions répulsives sont, par nature, de courte portée. On peut donc définir autour de chaque atome *i* une sphère de rayon r_{cut} finie contenant théoriquement toutes les informations nécessaires au calcul de E_i^{rep} . On appelle cet objet \mathbf{R}_i , constitué de la position et de l'élément d'un atome et de ses voisins dans un rayon, un **environnement atomique** (Fig. 5.9).



FIGURE 5.9 – Représentation schématique d'un environnement atomique.

L'expression du potentiel répulsif est définie comme une fonction de l'environnement atomique proche, c'est-à-dire comme un potentiel à plusieurs corps. Il s'agit d'un changement de perspective notable, les potentiels répulsifs de DFTB étant jusqu'à alors calculés comme une somme d'interactions de paires d'atomes.

En l'état, un environnement atomique constitue nos *données brutes*, à l'instar de l'image d'un chiffre manuscrit dans l'exemple *MNIST*. On rencontre là aussi certaines contraintes. Dans un premier temps, le nombre d'atomes présents dans l'environnement atomique peut varier; nous avons vu précédemment que la taille du vecteur d'entrée du réseau doit rester constante.

De plus, en envisageant de fournir les positions d'un atome et de ses voisins au réseau de neurones comme entrée, on se heurte à un souci de respect des invariants usuels en physique. L'énergie répulsive associée à un environnement atomique doit être :

• invariante par translation, c'est-à-dire que déplacer l'environnement dans l'espace ne doit pas modifier l'énergie répulsive associée. En pratique, si l'on déplace tous les atomes d'un système de 10 Å dans une direction \vec{d} , toutes les positions sont modifiées.

- invariante par rotation, c'est-à-dire que la rotation de l'environnement dans l'espace ne doit pas non plus modifier l'énergie répulsive associée. Évidemment, la rotation du système modifie les positions du système, ce qui est problématique dans notre cas.
- invariante par réflexion, c'est-à-dire que l'énergie répulsive d'un environnement doit être la même que celle de sa réflexion dans un miroir. Encore une fois, la transformation par réflexion modifie les positions des atomes.
- invariante par permutation d'atome équivalents, c'est-à-dire que la permutation de deux atomes de la même espèce chimique dans l'environnement ne doit pas modifier l'énergie répulsive du système. Dans les faits, deux neurones d'entrées d'un réseau ne sont pas équivalents et les permuter modifie la sortie du réseau.

La considération de ces invariants nous indique que les coordonnées brutes des atomes ne sont pas à même d'être utilisées en tant que vecteur d'entrée d'un réseau de neurones. Afin d'assurer que les potentiels répulsifs obtenus respectent tous ces impératifs, on choisit de les inclure dans notre prétraitement des données brutes : on associe donc, par constructions, les mêmes vecteurs d'entrée symétrisés aux environnements atomiques équivalents. Un tel vecteur symétrisé respectant ces invariants et contraintes, obtenu à partir d'un environnement atomique et fourni en entrée d'un réseau de neurones est appelé vecteur **descripteur**.

On se propose donc, dans cette section, de présenter la méthode des *fonctions de symétrie centrées sur les atomes*, utilisée dans cette thèse pour générer des descripteurs. Nous nous intéresserons aussi à quelques méthodes alternatives, utilisées dans la littérature.

5.4.1 Fonctions de symétrie centrées sur les atomes

Présentées pour la première fois par Jörg Behler en 2007 [77] puis affinées en 2011 [78], les fonctions de symétrie centrées sur les atomes (*Atom-Centered Symetry Functions* - **ACSF**) sont une classe de fonctions permettant de convertir un environnement atomique en vecteur descripteur. Ceux-ci sont convenablement symétrisés pour être utilisés comme vecteurs d'entrée d'une méthode d'apprentissage automatique - dans notre cas, des réseaux de neurones.

Tout d'abord, insistons sur le fait qu'un environnement atomique est défini dans un volume limité (Fig. 5.9); les contributions d'atomes plus distants que le rayon de coupure $r_{\rm cut}$ sont donc nulles.

Pour s'en assurer, on définit par rapport à la distance r à l'atome central une fonction de coupure $f_{cut}(r)$ dont la forme exacte peut varier. On l'utilise pour assurer une décroissance continue dans l'amplitude des contributions à mesure que l'on s'éloigne de l'atome central, jusqu'à s'annuler pour $r = r_{cut}$. La première définition de fonction de coupure présentée, analogue à celle utilisée dans quelques potentiels empiriques tels que le potentiel de Tersoff [79], est :

$$f_{\rm cut}(r) = \begin{cases} \frac{1}{2} \left(\cos\left(\frac{\pi r}{r_{\rm cut}}\right) + 1 \right) & \text{si} \quad 0 < r < r_{\rm cut}, \\ 0 & \text{sinon.} \end{cases}$$
(5.10)

Pour $r \ge r_{\text{cut}}$, cette expression de la fonction de coupure et sa dérivée sont nulles. Cependant, sa dérivée seconde présente une discontinuité au rayon de coupure; on lui préfère fréquemment une nouvelle forme, présentée quelques années plus tard.

$$f_{\rm cut}(r) = \begin{cases} \tanh^3 \left(1 - \frac{r}{r_{\rm cut}} \right) & \text{si} \quad 0 < r < r_{\rm cut}, \\ 0 & \text{sinon.} \end{cases}$$
(5.11)

Plusieurs formes de fonctions de symétrie sont initialement proposées mais l'usage en retient deux principales, majoritairement utilisées.

Les premières, que l'on nomme fonctions de symétrie radiales car elles ne dépendent explicitement que des distances interatomiques, sont des sommes de gaussiennes multipliées par la fonction de coupure. Elles s'expriment pour un atome i ainsi :

$$G_2^i = \sum_{j=1}^{N_{\text{vois}}^i} g_2(r_{ij}) = \sum_{j=1}^{N_{\text{vois}}^i} e^{-\eta(r_{ij}-r_s)^2} \cdot f_{\text{cut}}(r_{ij}).$$
(5.12)

On identifie trois paramètres définissant ces fonctions.

- η ajuste la largeur des gaussiennes; plus η est petit, plus la gaussienne est large.
- r_s est une distance de décalage à l'atome central; plus r_s est grand, plus le centre de la gaussienne est éloigné de l'atome central.
- $r_{\rm cut}$ est le rayon de coupure au-delà duquel les contributions atomiques sont nulles, présenté précédemment.

Chaque jeu de paramètres $(\eta, r_s, r_{\text{cut}})$ définit **une** fonction de symétrie radiale, dont on illustre quelques exemples dans la figure 5.10.

Les autres fonctions de symétrie principalement utilisées sont nommées fonctions de symétrie angulaires pour leur dépendance explicite aux distances inter-atomiques et aux angles θ_{ijk} formés par les triplets de l'environnement. Il s'agit là encore de gaussiennes multipliées par des fonctions de coupure, et incluant également une dépendance angulaire dans leur expression.

$$G_{4}^{i} = \sum_{j=1}^{N_{\text{vois}}^{i}} \sum_{\substack{k=1\\k\neq j}}^{N_{\text{vois}}^{i}} g_{4}\left(r_{ij}, r_{ik}, r_{jk}\right)$$

$$= 2^{1-\zeta} \sum_{j=1}^{N_{\text{vois}}^{i}} \sum_{\substack{k=1\\k\neq j}}^{N_{\text{vois}}^{i}}} (1 + \lambda \cos \theta_{ijk})^{\zeta} \cdot e^{-\eta r_{ij}^{2}} \cdot e^{-\eta r_{ik}^{2}} \cdot e^{-\eta r_{jk}^{2}} \cdot f_{\text{cut}}(r_{ij}) \cdot f_{\text{cut}}(r_{ik}) \cdot f_{\text{cut}}(r_{jk}).$$

(5.13)

Pour cette classe de fonctions de symétrie, on identifie quatre paramètres.



FIGURE 5.10 – Exemples de fonctions de symétrie G_2 .

- η qui détermine ici aussi la largeur des gaussiennes, de manière analogue aux fonctions de symétrie radiales (5.12).
- ζ fait office de résolution angulaire. Plus ζ est grand, plus la partie angulaire de la fonction de symétrie est étroite et donc sensible aux variations de θ_{ijk} .
- λ est un entier pouvant prendre les valeurs $\{-1; +1\}$, déplaçant respectivement le maximum du cosinus à $\theta_{ijk} = 0^{\circ}$ et $\theta_{ijk} = 180^{\circ}$.
- $r_{\rm cut}$ est le rayon de coupure au-delà duquel les contributions atomiques sont nulles, présenté précédemment.

Là encore, chaque jeu de paramètres $(\eta, \zeta, \lambda, r_{\text{cut}})$ définit **une** fonction de symétrie angulaire. Ces fonctions dépendant de trois distances inter-atomiques, leur visualisation est moins évidente que pour les fonctions de symétrie radiales. La partie radiale des fonctions G_4 est très similaire aux fonctions G_2 où $r_s = 0$, et leur partie angulaire est représentée dans la figure 5.11.

À partir de l'environnement associé à l'atome central i, on peut calculer G_2^i ou G_4^i pour différents jeux de paramètres; chaque fonction nous donne **un** élément du vecteur descripteur associé à l'environnement. En conséquence, un jeu de $N^{ACSF} = N_{G_2} + N_{G_4}$ fonctions de symétrie nous permet de construire un vecteur \vec{G} de taille N^{ACSF} . Ce vecteur respecte les invariants nécessaires à la construction d'un potentiel répulsif E_{rep} : translater un environnement atomique dans l'espace change les coordonnées absolues des atomes, mais ne modifie ni les distances ni les angles. Il en va de même pour toute opération de

Fonctions de symétrie angulaires - partie angulaire



FIGURE 5.11 – Partie angulaire de quelques fonctions de symétrie G_4 .

rotation ou de réflexion. De plus, la permutation d'atomes équivalents modifie simplement l'ordre des sommes des équations (5.12) et (5.13), sans en changer les valeurs finales; cet invariant est donc également respecté. Enfin, la taille du vecteur \vec{G} est constante, indépendamment du nombre d'atomes contenus dans l'environnement atomique.

Dès lors, nous pouvons donc utiliser le vecteur descripteur \vec{G} comme vecteur d'entrée pour un réseau de neurones ayant N^{ACSF} neurones dans sa couche d'entrée.

Générer des vecteurs descripteurs par cette approche présente plusieurs avantages. Tout d'abord, on utilise usuellement plusieurs dizaines de fonctions de symétrie par élément chimique : il y en a en général plus que le système n'a de degrés de liberté. On capture donc en principe la dimensionnalité du système, bien que les éventuelles redondances possibles puissent être problématiques pour un réseau de neurones. Par ailleurs il est possible de les choisir avec une intuition physique forte. Ainsi, pour un système donné, on peut construire des fonctions de symétrie très sensibles autour de distances ou d'angles inter-atomiques clés, très représentatifs du système. Enfin, chaque fonction de symétrie est indépendante des autres et l'on peut en ajouter ou en retirer à volonté; il est possible d'en générer une large quantité, et de sélectionner (voir annexe A) celles qui sont les plus pertinentes, par rapport au jeu de données de référence et/ou par rapport aux fonctions déjà retenues.

5.4.2 Autres descripteurs

Plusieurs alternatives aux fonctions de symétrie, visant à décrire l'environnement d'un atome en assurant de respecter les invariants nécessaires, sont fréquemment employées dans la littérature. On se propose d'en présenter deux ici : d'une part, la méthode du *bispectre de la densité des voisins*, classiquement utilisée en conjonction avec les méthodes à noyaux pour les potentiels à approximation gaussienne [71], et la méthode des *matrices de Coulomb* dont il existe certaines utilisations conjointes avec *DFTB* [44]. Ces deux approches sont, à l'inverse des fonctions de symétrie, complètement systématiques et requièrent un minimum d'intervention de l'utilisateur.

5.4.2.1 Bispectre de la densité des voisins

Intéressons nous tout d'abord à la méthode du bispectre de la densité des voisins. Cette approche s'affranchit de la sélection des paramètres vue pour les fonctions de symétrie. Pour cela, on commence par construire une densité de voisins en tout point de l'espace $n(\vec{r})$ pour tous les atomes i.

$$n(\vec{r}) = \delta\left(\vec{r}\right) + \sum_{r_{ij} < r_{\rm cut}} f_{\rm cut}\left(r_{ij}\right) \cdot \omega_j \cdot \delta\left(\vec{r} - \vec{r_{ij}}\right).$$

On utilise des fonctions δ non-nulles aux positions de l'atome *i* et de tous ses voisins *j* moins distants que le rayon de coupure r_{cut} .

$$\delta(x) = \begin{cases} 1 & \text{si} \quad x = 0, \\ 0 & \text{sinon.} \end{cases}$$

Afin d'assurer une décroissance continue de l'importance des contributions jusqu'à ce qu'elles s'annulent au-delà du rayon de coupure, on multiplie les contributions des voisins par une fonction de coupure analogue à celles utilisées pour les fonctions de symétrie (5.10) et (5.11). Enfin, les poids ω_j ne sont pas ici des paramètres mais des poids prédéfinis permettant de discriminer entre les éléments chimiques.

On projette la densité sur une sphère, obtenant une densité projetée $\hat{\rho}(\theta, \phi)$ qui rend compte de la distribution angulaire des voisins. On exprime cette densité projetée par un développement en harmoniques sphériques Y_{lm} :

$$\hat{\rho}(\theta,\phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} c_{lm} Y_{lm}(\theta,\phi) \, .$$

On représente ensuite la distribution radiale par un dernier angle additionnel θ_0 ; les trois angles (θ, ϕ, θ_0) définissent donc un jeu de positions sur la surface d'une 3-sphère³ de rayon $r_0 > r_{\text{cut}}$. La surface de cette 3-sphère est définie par l'ensemble des points $\mathbf{s} \in \mathbb{R}^4$ où $s_1^2 + s_2^2 + s_3^2 + s_4^2 = r_0^2$; les angles (θ, ϕ, θ_0) sont définis pour que :

$$s_{1} = r_{0} \cos \theta_{0}$$

$$s_{2} = r_{0} \sin \theta_{0} \cos \theta$$

$$s_{3} = r_{0} \sin \theta_{0} \sin \theta \cos \phi$$

$$s_{4} = r_{0} \sin \theta_{0} \sin \theta \sin \phi.$$

Les positions sont données par une projection, que l'on choisit [80] telle que :

^{3.} Sphère à quatre dimensions, aussi appelée hypersphère.

$$\vec{r} \to (\theta, \phi, \theta_0) = \left(\arctan\left(\frac{y}{x}\right), \arccos\left(\frac{z}{|\vec{r}|}\right), \frac{|\vec{r}|}{r_0} \right).$$

Si la projection d'un espace plat à trois dimension sur une sphère de dimension supérieure est difficile à visualiser, on peut s'en faire une intuition en s'intéressant à la *projection de Riemann* (Fig. 5.12), projetant un espace plat à une dimension (c'est-à-dire une droite) sur la surface d'une sphère à deux dimensions (c'est-à-dire un cercle).



FIGURE 5.12 – Projection de Riemann.

Tout comme nous l'avons vu pour la densité $\hat{\rho}(\theta, \phi)$ que l'on avait projeté sur une sphère avant de la développer en harmoniques sphériques, on développe la densité projetée sur une 3-sphère $\rho(\theta, \phi, \theta_0)$ en harmoniques hypersphériques $U_{m,m'}^j$.

$$\rho(\theta, \phi, \theta_0) = \sum_{j=1}^{\infty} \sum_{m,m'=-j}^{j} c_{m,m'}^{j} U_{m,m'}^{j}.$$

On utilise les coefficients $c_{m,m'}^j = \langle U_{m,m'}^j | \rho \rangle$ pour construire notre descripteur, la *matrice du bispectre*, dont les éléments sont définis comme

$$B_{j_1,j_2,j} = \sum_{m_1,m_1'=-j_1}^{j_1} \sum_{m_2,m_2'=-j_2}^{j_2} \sum_{m,m'=-j}^{j} \left(c_{m,m'}^j \right) \cdot C_{j_1m_1j_2m_2}^{jm} \times c_{j_1m_1'j_2m_2'}^{jm'} \cdot c_{m_1',m_1}^{j_1} \cdot c_{m_2',m_2}^{j_2},$$

où les coefficients $C_{j_1m_1j_2m_2}^{jm}$ sont appelés coefficients de *Clebsch-Gordan*, rencontrés initialement en théorie des invariants et apparaissant également en mécanique quantique lors de l'étude des couplages de moments angulaires [81]. Cette matrice est à la base utilisée en reconnaissance d'images et de motifs [82]; elle respecte aussi tous les invariants requis pour un vecteur descripteur. La capacité à augmenter à volonté le nombre d'harmoniques hypersphériques - au prix de temps de calcul - pour affiner la représentation de l'environnement est un avantage certain. En revanche, l'utilisation de fonctions δ implique des discontinuités dans la représentation des environnements : une modification mineure d'une position atomique impacte fortement le descripteur. En conséquence, la majorité des applications actuelles remplacent ces fonctions δ par des gaussiennes [71,80].

5.4.2.2 Matrices de Coulomb

Enfin, intéressons nous à une dernière approche visant à générer des vecteurs descripteurs : la méthode des matrices de Coulomb. Initialement proposées par M. Rupp et al. [83] en 2012, elles sont notamment utilisées pour générer des corrections aux potentiels répulsifs $E_{\rm rep}$ pour DFTB [44] et trouvent leurs principales applications dans l'étude de systèmes moléculaires. On cherche avec cet outil à décrire l'environnement d'une liaison entre atomes, et non pas l'environnement atomique (Fig. 5.13).



FIGURE 5.13 – Concept d'environnement de liaison atomique.

On identifie une liaison atomique entre des atomes A et B, d'espèce chimique quelconque t(A) et t(B), si la distance entre eux r_{AB} est inférieure au rayon de coupure $r_{\text{cut}}^{t(A)t(B)} = \min\left(r_{\text{cut}}^{t(a)}, r_{\text{cut}}^{t(b)}\right)$, qui dépend des espèces chimiques de A et de B. On inclut dans l'environnement de liaison tout atome O_i , d'espèce chimique quelconque $t(O_i)$, étant voisin de A ou B.

$$\min_{C=A,B} |R_{O_i} - R_C| < r_{\mathrm{cut}}^{t(c)}$$

La définition de cet environnement de liaison nous permet de construire la matrice de

Coulomb associée.

$$b(A,B) = \begin{pmatrix} 0.5\eta Z_{t(A)}^{2.4} & \frac{Z_{t(A)}Z_{t(B)}}{|R_A - R_B|} & \frac{Z_{t(A)}Z_{t(O_1)}}{|R_A - R_{O_1}|} & \cdots \\ \frac{Z_{t(B)}Z_{t(A)}}{|R_B - R_A|} & 0.5\eta Z_{t(B)}^{2.4} & \frac{Z_{t(B)}Z_{t(O_1)}}{|R_B - R_{O_1}|} & \cdots \\ \frac{Z_{t(O_1)}Z_{t(A)}}{|R_{O_1} - R_A|} & \frac{Z_{t(O_1)}Z_{t(B)}}{|R_{O_1} - R_B|} & 0.5\eta Z_{t(O_1)}^{2.4} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Ici, le poids associé sur la diagonale $\eta > 1$ permet de renforcer la différenciation entre atomes de types différents. La forme des éléments de matrice en 1/R est très semblable à celle des interactions coulombiennes, et est à l'origine du nom de la méthode; elle permet aussi et surtout d'amoindrir la contribution des atomes les plus éloignés. L'absence formelle de *fonction de coupure* implique néanmoins une variation discontinue des contributions autour du rayon de coupure r_{cut} .

Afin d'assurer que le descripteur respecte l'invariance par permutation, les lignes de la matrice sont triées par normes décroissantes :

$$\left(\frac{1}{2}\eta Z_{t(A)}^{2,4}\right)^{2} + \sum_{C \neq A} \left(\frac{Z_{t(A)}Z_{t(C)}}{|R_{A} - R_{C}|}\right)^{2} \ge \left(\frac{1}{2}\eta Z_{t(B)}^{2,4}\right)^{2} + \sum_{C \neq B} \left(\frac{Z_{t(B)}Z_{t(C)}}{|R_{B} - R_{C}|}\right)^{2} \ge \left(\frac{1}{2}\eta Z_{t(O_{1})}^{2,4}\right)^{2} + \sum_{C \neq O_{1}} \left(\frac{Z_{t(O_{1})}Z_{t(C)}}{|R_{O_{1}} - R_{C}|}\right)^{2} \ge \cdots$$

Une fois la matrice ordonnée, on la transforme en vecteur en l'*aplatissant*, c'est-à-dire en sélectionnant ses éléments ligne par ligne.

L'un des principaux défauts de cette méthode est évident : la taille de la matrice est dépendante du nombre d'atomes compris dans l'environnement de la liaison atomique. Les auteurs de la méthode proposent de contourner ce problème en choisissant une taille maximale - c'est-à-dire un nombre maximal de voisins possible N^{\max} . Cette solution, relativement inélégante, pose quelques problèmes d'efficacité numérique dans le cas où il y a beaucoup moins que N^{\max} voisins, auquel cas on manipule des objets inutilement volumineux car remplis de zéros, mais aussi et surtout dans le cas où il y a plus que N^{\max} voisins.

Pour la suite, nous utilisons uniquement des fonctions de symétrie centrées sur les atomes : G_2 et G_4 .

Chapitre 6

Application aux potentiels répulsifs

L'objectif de cette thèse est de permettre au code de simulation atomistique semiempirique DFTB+ d'améliorer sa représentation des interactions répulsives entre atomes grâce aux outils d'apprentissage automatique.

La présentation et l'étude des outils à notre disposition dans la section 5.3 justifient notre choix d'utiliser des réseaux de neurones à propagation avant.

Le fonctionnement des réseaux de neurones peut être séparé en deux processus distincts (voir 5.3.1) : l'apprentissage et le calcul. Ce chapitre s'intéresse à l'application de chacun d'eux aux travaux de cette thèse : c'est-à-dire le détail du calcul de l'énergie répulsive (et des forces répulsives en annexe B), mais également la présentation de quelques algorithmes d'optimisation.

6.1 Calcul de l'énergie répulsive

Nous avons explicité précédemment le processus de calcul d'un réseau de neurones (voir 5.3.1) et le calcul de vecteurs descripteurs avec les fonctions de symétrie centrées sur les atomes (voir 5.4.1). Dans cette section, nous détaillons comment utiliser ces outils pour calculer l'énergie répulsive totale du système. Le calcul des forces répulsives, très analogue mais assez volumineux, est quant à lui donné en annexe B.

Nos outils sont construits pour calculer le potentiel répulsif associé à chaque atome : connaissant l'environnement atomique d'**un** atome *i*, on détermine le vecteur descripteur associé que l'on fournit au réseau de neurones pour obtenir l'énergie répulsive appliquée sur **cet** atome $E_i^{\text{rep,NN}}$.

Une représentation schématique du calcul de cette grandeur est donnée dans la figure 6.1. Afin de définir précisément ce processus, adoptons les notations suivantes :

- \mathbf{R}_i désigne l'environnement atomique de l'atome *i*, c'est-à-dire sa position $\vec{r_i}$ et $\{\vec{r_j}\}$ celles de tous ses voisins *j*.
- $\vec{G}(\mathbf{R}_{\mathbf{i}})$ désigne le vecteur descripteur associé à cet environnement.

• NN $\begin{bmatrix} \vec{X} \end{bmatrix}$ désigne le retour d'un réseau de neurones ayant pour vecteur d'entrée \vec{X} , comme défini précédemment (5.4).

On peut dès lors exprimer l'énergie répulsive calculée par un réseau de neurones pour un environnement atomique \mathbf{R}_{i} .



$$E_{\rm rep}^{i,\rm NN} = \rm NN}\left[\vec{G}\left(\mathbf{R_i}\right)\right]. \tag{6.1}$$

FIGURE 6.1 – Calcul de l'énergie répulsive d'un environnement atomique.

Dans le cas où l'on aurait accès directement aux énergies répulsives atomiques de référence $E_{\text{rep},i}^{\text{ref}}$, l'application serait très standard et correspondrait à un cas d'école d'utilisation de réseaux de neurones.

Pour rappel, l'énergie répulsive du jeu de données de référence est calculée comme un terme correctif : on connaît l'énergie totale DFT associée au système de référence $E_{\text{DFT}}^{\text{ref}}$, et on détermine l'énergie totale DFTB sans contribution répulsive $E_{\text{DFTB},\mathcal{E}_{\text{rep}}}^{\text{ref}}$. Les méthodes DFT et DFTB ne fournissent pas d'information énergétique à l'échelle des atomes, et nous n'avons accès comme valeur de référence qu'à l'énergie répulsive totale du système. Celle-ci représente la somme des énergies répulsives de chaque environnement contenu dans la boîte de simulation, et est calculé comme $E_{\text{rep.boite}}^{\text{ref}}$:

$$E_{\rm rep,boite}^{\rm ref} = E_{\rm DFT}^{\rm ref} - E_{\rm DFTB, \underline{E}_{\rm rep}}^{\rm ref}.$$
(6.2)

Comme présenté dans la figure 6.1, notre méthodologie place chaque réseau de neurones à l'échelle d'un environnement atomique, retournant donc l'énergie répulsive exercée sur un seul atome i, $E_{\text{rep},i}^{\text{NN}}$. Néanmoins, nous n'avons pas directement accès à l'énergie répulsive de référence $E_{\text{rep},i}^{\text{ref}}$.

Il est donc nécessaire de reproduire avec nos outils l'énergie répulsive de la boîte de simulation entière $E_{\text{rep,boite}}^{\text{NN}}$ (Fig. 6.2).

$$E_{\rm rep,boite}^{\rm ref} \approx E_{\rm rep,boite}^{\rm NN} = \sum_{i=1}^{N_{\rm atomes}^{\rm boite}} E_{{\rm rep},i}^{\rm NN}.$$
(6.3)

Avec cette approche, on obtient la grandeur recherchée en sommant les valeurs de sortie de plusieurs réseaux de neurones atomiques. Au lieu d'utiliser le réseau de neurones pour prédire directement l'énergie répulsive de la boîte de simulation entière, on introduit une granularité supplémentaire en utilisant plusieurs copies d'un même réseau, chacune chargée de rendre compte du comportement d'une sous-partie du système. Dans la litté-rature, les potentiels employant cette approche sont nommés potentiels numériques par réseaux de neurones de haute dimension [84].

Insistons une fois de plus sur un point crucial, déjà évoqué plus haut (6.3) : il n'est pas nécessaire de connaître l'énergie répulsive par atome de référence $E_{\text{rep},i}^{\text{ref}}$ pour entraîner les réseaux de neurones atomiques à restituer $E_{\text{rep},i}^{\text{NN}}$. Nous avons simplement besoin d'avoir accès aux énergies répulsives de boîtes $E_{\text{rep},\text{boite}}^{\text{ref}}$; une explication détaillée est présentée plus loin (voir 6.2).



FIGURE 6.2 – Calcul de l'énergie répulsive d'une boîte de simulation.

6.2 Optimiser un réseau de neurones

L'utilité d'un réseau de neurones est évidemment dépendante de qualité de représentation : plus il est capable de reproduire de comportements du jeu de données de référence, plus il est à même de trouver une application. La problématique de l'entraînement d'un réseau, c'est-à-dire comment lui apprendre à rendre compte d'un jeu de données, est la
question centrale de toute application ayant recours à ce genre d'outils. De nombreux algorithmes existent, certains plus complexes, populaires ou originaux que les autres.

Durant la première partie de ces travaux de thèse, nous avons développé une librairie *Fortran* (dont le détail est disponible en annexe C.1.2.1) implémentant les concepts essentiels à notre application (environnement atomique, boîte de simulation, réseaux de neurones de haute dimension, fonctions de symétrie centrées sur les atomes), avec pour vocation finale d'être greffée à un code de calcul *DFTB*. En amont de cela, sa fonction initiale était de tester les différents outils conceptuels en jeu dans le formalisme présenté précédemment, en construisant et entraînant des réseaux de neurones sur des systèmes simples pour mieux appréhender l'importance de chaque composante.

L'algorithme initialement utilisé pour optimiser ces réseaux de neurones, la rétropropagation du gradient, est présenté ci-après (voir 6.2.1). L'implémentation de cet algorithme a évolué graduellement, en fonction des besoins : initialement concentré sur les énergies par atome issues de calculs de dynamique moléculaire classique, il fut étendu aux énergies totales de boîtes et aux énergies répulsives extraites de calculs premier principe (voir 6.2.1.1). Les performances obtenues étaient prometteuses mais se heurtaient à une limitation de taille : on ne considérait que des énergies de référence à l'échelle de la boîte de simulation entière, et le manque d'informations locales semblait plafonner la précision accessible. Nous avons donc cherché à introduire le concept de forces répulsives appliquées sur chacun des atomes au sein de la librairie.

La mise en place du calcul des forces (voir annexe B), et *a fortiori* leur prise en compte lors du processus d'optimisation se sont révélés très chronophages. Après plusieurs mois d'essai-erreur, de débogage et de validation numérique, nous sommes parvenus à étendre l'algorithme de rétropropagation pour pouvoir optimiser conjointement sur l'énergie et les forces répulsives (voir 6.2.1.2).

Les précisions accessibles restaient toutefois sensiblement insuffisantes, et nous avons finalement choisi de remettre nos entraînements de réseaux de neurones dans les mains d'un logiciel externe, N2P2, proposant un algorithme d'optimisation plus exotique : le filtre de Kalman étendu, présenté plus loin (voir 6.2.2). Celui-ci permet, dans le cas de petits réseaux tels que ceux que l'on manipule dans les présents travaux, un apprentissage robuste. S'il ne permet pas un gain d'un ordre de grandeur en terme d'erreur, il offre tout de même de gagner au moins un facteur deux sur les précisions obtenues.

6.2.1 Rétropropagation du gradient

Comme tout algorithme d'optimisation, l'idée très générale de la rétropropagation du gradient est d'ajuster les paramètres d'une approximation - ici les poids et biais d'un réseau de neurones - afin de *reproduire convenablement* l'information contenue dans un jeu de données de référence. Afin d'évaluer la qualité de reproduction d'un réseau de neurones pour un jeu donné, on définit une fonction coût Γ , a priori quelconque mais qui doit être inversement proportionnelle à la qualité de représentation. Si le réseau améliore sa représentation, la fonction coût diminue; il s'agit donc d'une forme de quantification de l'erreur. Dans notre cas, on peut imaginer plusieurs fonctions Γ_E , Γ_F ou Γ_{EF} selon

que l'on évalue la capacité à reproduire individuellement ou conjointement les énergies et forces répulsives du jeu de données de référence.

Si l'on veut évaluer la qualité de représentation des énergies répulsives du jeu de référence $\left\{E_{\text{rep},i}^{\text{ref}}\right\}_{i=1}^{N^{\text{boites}}}$, une proposition pour Γ_E s'écrit :

$$\Gamma_E = \frac{1}{2N^{\text{boites}}} \sum_{i=1}^{N^{\text{boites}}} \left(E_{\text{rep},i}^{\text{NN}} - E_{\text{rep},i}^{\text{ref}} \right)^2.$$
(6.4)

De même, on peut définir une fonction coût Γ_F permettant d'évaluer la capacité de représentation des forces répulsives.

$$\Gamma_F = \frac{1}{2N^{\text{boites}}} \sum_{i=1}^{N^{\text{boites}}} \frac{1}{3N^i_{\text{atomes}}} \sum_{j=1}^{N^i_{\text{atomes}}} \sum_{\alpha=\{x,y,z\}} \left(F^{\text{rep},\text{NN},i}_{j,\alpha} - F^{\text{rep},\text{ref},i}_{j,\alpha}\right)^2.$$
(6.5)

Enfin, on peut définir une fonction coût hybride Γ_{EF} , dans laquelle on attribue un poids relatif β permettant d'ajuster l'importance relative des énergies et forces répulsives.

$$\Gamma_{EF} = \Gamma_E + \beta \cdot \Gamma_F \tag{6.6}$$

Une fois la fonction coût choisie, l'idée de l'algorithme de rétropropagation du gradient est de faire évoluer itérativement tous les poids w et biais b du réseau de manière à minimiser la fonction Γ . Pour cela, on propage l'erreur depuis la couche de sortie du réseau vers sa couche d'entrée, et on évalue à chaque couche le gradient de la fonction coût en fonction des paramètres. En modifiant les paramètres dans la direction opposée au gradient, on diminue en principe la fonction coût.

$$w^{t+1} = w^t - \lambda \frac{\partial \Gamma}{\partial w^t}$$
 et $b^{t+1} = b^t - \lambda \frac{\partial \Gamma}{\partial b^t}$. (6.7)

On remarque la présence du facteur λ , ou *learning rate*¹ qui définit l'amplitude de l'actualisation des paramètres. Sa valeur est déterminante pour l'apprentissage, influant à la fois sur la vitesse de convergence et sur la robustesse du processus. Le choix de cette valeur est cependant hautement non trivial (Fig. 6.3), dépend fortement du système considéré et revient à l'utilisateur.

Si λ est grand, chaque époque² d'apprentissage modifie beaucoup la valeur des paramètres : on se rapproche rapidement du minimum, mais on peut ne jamais l'atteindre si la fonction à représenter fait état de variations trop fines. À l'inverse, si λ est petit, l'optimisation est robuste et l'on s'assure d'atteindre un minimum, au prix d'une longue optimisation; on peut également se trouver bloqué dans un minimum local, avec des pas trop petits pour en sortir. Évidemment, il faut trouver un équilibre pour que λ ne soit ni trop grand ni trop petit, même si ces qualificatifs sont impossibles à quantifier en général et dépendent du problème considéré.

^{1.} Parfois francisé en «taux d'apprentissage».

^{2.} Itération du processus durant laquelle on parcourt l'intégralité du jeu de données d'apprentissage.



FIGURE 6.3 – Schématisation de l'impact du taux d'apprentissage

L'optimisation (6.7) requiert donc d'évaluer l'influence individuelle de chacun des poids et biais sur la valeur de la fonction coût, celle-ci pouvant inclure ou non énergies et forces répulsives.

6.2.1.1 Apprentissage sur l'énergie répulsive

Dans le cas où l'on souhaite entraîner un réseau de neurones à reproduire les énergies répulsives (6.4), il est nécessaire de déterminer :

$$\frac{\partial \Gamma_E}{\partial w} = \frac{\partial}{\partial w} \left[\frac{1}{2N^{\text{boites}}} \sum_{i=1}^{N^{\text{boites}}} \left(E_{\text{rep},i}^{\text{NN}} - E_{\text{rep},i}^{\text{ref}} \right)^2 \right] \\
= \frac{1}{N^{\text{boites}}} \sum_{i=1}^{N^{\text{boites}}} \left(E_{\text{rep},i}^{\text{NN}} - E_{\text{rep},i}^{\text{ref}} \right) \cdot \frac{\partial E_{\text{rep},i}^{\text{NN}}}{\partial w}.$$
(6.8)

avec (6.3)

$$\frac{\partial E_{\mathrm{rep},i}^{\mathrm{NN}}}{\partial w} = \sum_{j=1}^{N_{\mathrm{atomes}}^{i}} \frac{\partial E_{\mathrm{rep},i,j}^{\mathrm{NN}}}{\partial w},$$

d'où

$$\frac{\partial \Gamma_E}{\partial w} = \frac{1}{N^{\text{boites}}} \sum_{i=1}^{N^{\text{boites}}} \left(E_{\text{rep},i}^{\text{NN}} - E_{\text{rep},i}^{\text{ref}} \right) \cdot \left(\sum_{j=1}^{N_{\text{atomes}}^i} \frac{\partial E_{\text{rep},i,j}^{NN}}{\partial w} \right), \tag{6.9}$$

et similairement, pour les biais

$$\frac{\partial \Gamma_E}{\partial b} = \frac{1}{N^{\text{boites}}} \sum_{i=1}^{N^{\text{boites}}} \left(E_{\text{rep},i}^{\text{NN}} - E_{\text{rep},i}^{\text{ref}} \right) \cdot \left(\sum_{j=1}^{N^{i}_{\text{atomes}}} \frac{\partial E_{\text{rep},i,j}^{\text{NN}}}{\partial b} \right).$$
(6.10)

En nommant $\vec{A}^l = \mathbf{w}^l \cdot \vec{X}^l + \vec{b}^l$ le vecteur d'activation de la couche l, on a pour un réseau à L couches cachées :

$$\begin{split} E_{\text{rep},i,j}^{\text{NN}} &= f^{L+1} \left(\vec{A}^{L+1} \right) = f^{L+1} \left(\mathbf{w}^{\mathbf{L}+\mathbf{1}} \cdot \vec{X}^{L+1} + \vec{b}^{L+1} \right) = f^{L+1} \left(\mathbf{w}^{\mathbf{L}+\mathbf{1}} \cdot \vec{Y}^{L} + \vec{b}^{L+1} \right) \\ \vec{Y}^{L} &= f^{L} \left(\vec{A}^{L} \right) = f^{L} \left(\mathbf{w}^{\mathbf{L}} \cdot \vec{X}^{L} + \vec{b}^{L} \right) = f^{L} \left(\mathbf{w}^{\mathbf{L}} \cdot \vec{Y}^{L-1} + \vec{b}^{L} \right) \\ \vdots \\ \vec{Y}^{1} &= f^{1} \left(\vec{A}^{1} \right) = f^{1} \left(\mathbf{w}^{\mathbf{1}} \cdot \vec{X}^{1} + \vec{b}^{1} \right) = f^{1} \left(\mathbf{w}^{\mathbf{1}} \cdot \vec{Y}^{0} + \vec{b}^{1} \right) = f^{1} \left(\mathbf{w}^{\mathbf{1}} \cdot \vec{G}_{i,j} + \vec{b}^{1} \right). \end{split}$$

Explicitions la dérivée $\frac{\partial E_{\text{rep},i,j}^{\text{NN}}}{\partial w_{n_1,n_2}^l}$ par rapport au poids liant le neurone n_1 de la couche l et le neurone n_2 de la couche l-1. Proposons une approche par couches successives, où l'on s'intéresse à l'influence du poids sur le vecteur de sortie d'une couche l', $\frac{\partial \vec{Y}^{l'}}{\partial w_{n_1,n_2}^l}$. On peut définir le vecteur de sortie $\vec{Y}^{l'}$ comme

$$\frac{\partial \vec{Y}^{l'}}{\partial w_{n_1,n_2}^l} = \frac{\partial}{\partial w_{n_1,n_2}^l} \left[f^{l'} \left(\vec{A}^{l'} \right) \right] = \frac{\partial}{\partial w_{n_1,n_2}^l} \left[f^{l'} \left(\mathbf{w}^{\mathbf{l}'} \cdot \vec{X}^{l'} + \vec{b}^{l'} \right) \right]$$
$$= \frac{\partial f^{l'} \left(\vec{A}^{l'} \right)}{\partial \vec{A}^{l'}} \cdot \frac{\partial}{\partial w_{n_1,n_2}^l} \left[\mathbf{w}^{\mathbf{l}'} \cdot \vec{X}^{l'} + \vec{b}^{l'} \right].$$

Le premier terme, c'est-à-dire la dérivée de la fonction de transfert par rapport à l'activation, est calculé en annexe (voir (B.9), (B.10) et (B.11)).

• Dans le cas où l'on s'intéresse à la couche l' < l, la modification d'un poids placé plus loin dans le réseau n'influe pas sur le vecteur de sortie car on s'intéresse exclusivement à des réseaux de neurones multicouches à propagation avant.

$$\frac{\partial \vec{Y}^{l'}}{\partial w_{n_1,n_2}^l} = 0 \quad \text{si} \quad l' < l. \tag{6.11}$$

• En considérant le cas particulier où l' = l, la modification d'un poids w_{n_1,n_2}^l impacte évidemment $\vec{Y}^{l'}$. Étudions les éléments de $\frac{\partial \vec{A}^l}{\partial w_{n_1,n_2}^l}$ neurone par neurone. On rappelle que :

$$A_n^l = b_n^l + \sum_{n'=1}^{N_n^{l-1}} w_{n,n'}^l \cdot X_{n'}^l.$$

On distingue donc deux cas :

$$\frac{\partial A_n^l}{\partial w_{n_1,n_2}^l} = \begin{cases} X_{n_2}^l & \text{si } n = n_1 \\ 0 & \text{sinon} \end{cases}$$
(6.12)

On a donc, au final

$$\frac{\partial Y_n^l}{\partial w_{n_1,n_2}^l} = \begin{cases} \frac{\partial f^l(A_{n_1}^l)}{\partial A_{n_1}^l} \cdot X_{n_2}^l & \text{si } n = n_1 \\ 0 & \text{sinon} \end{cases}$$
(6.13)

• Enfin, si l'on considère le cas où l' > l, la modification du poids w_{n_1,n_2}^l affecte le vecteur de sortie de la couche l, \vec{Y}^l , et donc les vecteurs d'entrées - et de sortie - de toutes les couches postérieures. On a

$$\frac{\partial A_n^{l'}}{\partial w_{n_1,n_2}^l} = \frac{\partial}{\partial w_{n_1,n_2}^l} \left[\vec{b}_n^{l'} + \sum_{n'=1}^{N_n^{l'-1}} w_{n,n'}^{l'} \cdot X_{n'}^{l'} \right] = \sum_{n'=1}^{N_n^{l'-1}} \frac{\partial Y_{n'}^{l'-1}}{\partial w_{n_1,n_2}^l} \cdot w_{n,n'}^{l'},$$

d'où

• Cas où l' = l

• Cas où l' > l

$$\frac{\partial Y_n^{l'}}{\partial w_{n_1,n_2}^l} = \frac{\partial f^{l'}\left(A_n^{l'}\right)}{\partial A_n^{l'}} \cdot \sum_{n'=1}^{N_n^{l'-1}} \frac{\partial Y_{n'}^{l'-1}}{\partial w_{n_1,n_2}^l} \cdot w_{n,n'}^{l'} \quad \text{si} \quad l' > l.$$
(6.14)

Cette démarche nous permet de déterminer la dérivée de $E_{\text{rep},i,j}^{\text{NN}}$ (retournée par la couche de sortie du réseau) par rapport à n'importe quel poids w_{n_1,n_2}^l . On peut similairement déterminer ces mêmes dérivées, cette fois-ci par rapport à n'importe quel biais du réseau $b_{n_1}^l$ associé au neurone n_1 de la couche l.

- Cas où l' < l $\frac{\partial \vec{Y}^{l'}}{\partial b_{n_1}^l} = 0.$ (6.15)
 - $\frac{\partial Y_n^l}{\partial b_{n_1}^l} = \begin{cases} \frac{\partial f^l(A_{n_1}^l)}{\partial A_{n_1}^l} & \text{si } n = n_1 \\ 0 & \text{sinon} \end{cases}$ (6.16)
 - $\frac{\partial Y_n^{l'}}{\partial b_{n_1}^l} = \frac{\partial f^{l'}\left(A_n^{l'}\right)}{\partial A_n^{l'}} \cdot \sum_{n'=1}^{N_n^{l'-1}} \frac{\partial Y_{n'}^{l'-1}}{\partial b_{n_1}^l} \cdot w_{n,n'}^{l'}.$ (6.17)

6.2.1.2 Apprentissage sur les forces répulsives

Lorsque l'on souhaite prendre en compte les forces répulsives dans l'optimisation, que ce soit individuellement (6.5) ou conjointement avec les énergies répulsives (6.6), il est nécessaire d'évaluer $\frac{\partial \Gamma_F}{\partial w}$ et $\frac{\partial \Gamma_F}{\partial b}$.

$$\frac{\partial \Gamma_F}{\partial w} = \frac{\partial}{\partial w} \left[\frac{1}{2N^{\text{boites}}} \sum_{i=1}^{N^{\text{boites}}} \frac{1}{3N_{\text{atomes}}^i} \sum_{j=1}^{N^{\text{atomes}}} \sum_{\alpha = \{x,y,z\}}^{N^{\text{atomes}}} \left(F_{j,\alpha}^{\text{rep,NN},i} - F_{j,\alpha}^{\text{rep,ref},i}\right)^2 \right]$$
$$= \frac{1}{2N^{\text{boites}}} \sum_{i=1}^{N^{\text{boites}}} \frac{1}{3N_{\text{atomes}}^i} \sum_{j=1}^{N^{\text{atomes}}} \sum_{\alpha = \{x,y,z\}} \frac{\partial}{\partial w} \left[\left(F_{j,\alpha}^{\text{rep,NN},i} - F_{j,\alpha}^{\text{rep,ref},i}\right)^2 \right]$$

$$=\frac{1}{N^{\text{boites}}}\sum_{i=1}^{N^{\text{boites}}}\frac{1}{3N_{\text{atomes}}^{i}}\sum_{j=1}^{N_{\text{atomes}}^{i}}\sum_{\alpha=\{x,y,z\}}\left(F_{j,\alpha}^{\text{rep,NN},i}-F_{j,\alpha}^{\text{rep,ref},i}\right)\frac{\partial F_{j,\alpha}^{\text{rep,NN},i}}{\partial w}$$

On montre en annexe (voir B) que :

$$F_{j,\alpha}^{\mathrm{rep,NN},i} = -\sum_{k=1}^{N_{\mathrm{atomes}}^{i}} J\left(T^{L+1}, \vec{X}^{L+1}\right) \cdot J\left(T^{L}, \vec{X}^{L}\right) \cdot \cdots \cdot J\left(T^{0}, \vec{G}_{ik}\right) \cdot \frac{\partial G_{ik}}{\partial r_{j,\alpha}^{i}},$$

où $T^l(\vec{X^l})=\vec{Y}^l$ désigne la transformation effectuée par la $l^{\rm ème}$ couche du réseau. Cela donne :

$$\frac{\partial F_{j,\alpha}^{\text{rep,NN},i}}{\partial w_{n_1,n_2}^l} = -\sum_{k=1}^{N_{\text{atomes}}^i} \frac{\partial}{\partial w_{n_1,n_2}^l} \left[J\left(T^{L+1}, \vec{X}^{L+1}\right) \cdot J\left(T^L, \vec{X}^L\right) \cdot \cdots \right. \\ \left. \cdot J\left(T^l, \vec{X}^l\right) \cdot \cdots \cdot J\left(T^0, \vec{G}_{ik}\right) \right] \cdot \frac{\partial G_{ik}}{\partial r_{j,\alpha}^i}.$$

Nous détaillons également en annexe le calcul des dérivées des fonctions de symétrie (voir B.2); on donne (B.12) également l'expression des jacobiens des transformations effectuées dans chaque couche l > 0.

$$J\left(T^{l}, \vec{X}^{l}\right) = \begin{bmatrix} w_{1,1}^{l} \frac{\partial f^{l}\left(A_{1}^{l}\right)}{\partial A_{1}^{l}} & \cdots & w_{1,N_{n}^{l}}^{l} \frac{\partial f^{l}\left(A_{1}^{l}\right)}{\partial A_{1}^{l}} \\ \vdots & \ddots & \vdots \\ w_{N_{n}^{l-1},1}^{l} \frac{\partial f^{l}\left(A_{N_{n}^{l}}^{l}\right)}{\partial A_{N_{n}^{l}}^{l}} & \cdots & w_{N_{n}^{l-1},N_{n}^{l}}^{l} \frac{\partial f^{l}\left(A_{N_{n}^{l}}^{l}\right)}{\partial A_{N_{n}^{l}}^{l}}. \end{bmatrix}$$

Lorsque l'on souhaite calculer :

$$\frac{\partial}{\partial w_{n_1,n_2}^l} \left[J\left(T^{L+1}, \vec{X}^{L+1}\right) \cdot J\left(T^L, \vec{X}^L\right) \cdot \cdots \cdot J\left(T^l, \vec{X}^l\right) \cdot \cdots \cdot J\left(T^0, \vec{G}_{ik}\right) \right].$$

il faut tout d'abord identifier les jacobiens dépendant de w_{n_1,n_2}^l :

- Les jacobiens des couches précédant la couche l ne dépendent pas de w_{n_1,n_2}^l , car on considère uniquement des réseaux de neurones multicouches à propagation avant ; il n'y a aucune rétroaction.
- Dans le cas où l'on considère le jacobien de la couche *l*, la dépendance est explicite et la dérivée est non-nulle.
- Tout comme dans la section précédente (voir 6.2.1.1), la propagation des valeurs dans le réseau implique que la modification d'un poids w_{n_1,n_2}^l dans la couche l affecte les vecteurs d'entrées, et donc les jacobiens, de toutes les couches suivantes.

On peut résumer :

$$\frac{\partial J\left(T^{l\prime}, \vec{X}^{l\prime}\right)}{\partial w_{n_1, n_2}^l} \begin{cases} = 0 & \text{si } l' < l \\ \neq 0 & \text{si } l' = l \\ \neq 0 & \text{si } l' > l \end{cases}.$$

Ce qui nous laisse réécrire :

$$\frac{\partial F_{j,\alpha}^{\text{rep,NN},i}}{\partial w_{n_1,n_2}^l} = -\sum_{k=1}^{N_{\text{atomes}}^i} \frac{\partial}{\partial w_{n_1,n_2}^l} \left[J\left(T^{L+1}, \vec{X}^{L+1}\right) \cdot J\left(T^L, \vec{X}^L\right) \cdot \cdots \right. \\ \left. \cdot J\left(T^l, \vec{X}^l\right) \right] \cdot \cdots \cdot J\left(T^0, \vec{G}_{ik}\right) \cdot \frac{\partial G_{ik}}{\partial r_{j,\alpha}^i}.$$

Intéressons nous tout d'abord au cas où l'on considère la couche l, contenant le poids w_{n_1,n_2}^l . Si l'on adopte la notation $J^{a,b}\left(T^l, \vec{X^l}\right) = w_{a,b}^l \frac{f^l(A_a^l)}{\partial A_a^l}$ pour désigner les éléments des matrices jacobiennes, on peut écrire :

$$\frac{\partial J^{a,b}\left(T^{l},\vec{X^{l}}\right)}{\partial w_{n_{1},n_{2}}^{l}} = \begin{cases} 0 & \text{si } a \neq n_{1} \\ \frac{f^{l}\left(A_{a}^{l}\right)}{\partial A_{a}^{l}} + w_{a,b}^{l} \frac{\partial^{2} f^{l}\left(A_{a}^{l}\right)}{\partial w_{n_{1},n_{2}}^{l} \partial A_{a}^{l}} & \text{si } a = n_{1} \text{ et } b = n_{2} \\ w_{a,b}^{l} \frac{\partial^{2} f^{l}\left(A_{a}^{l}\right)}{\partial w_{n_{1},n_{2}}^{l} \partial A_{a}^{l}} & \text{si } a = n_{1} \text{ et } b \neq n_{2} \end{cases}$$

On remarque ici que l'unique terme inconnu est la dérivée d'ordre 2 des fonctions de transfert. En utilisant le théorème de dérivation des fonctions composées, on peut réécrire :

$$\frac{\partial^2 f^l \left(A_a^l \right)}{\partial w_{n_1,n_2}^l \partial A_a^l} = \frac{\partial}{\partial A_a^l} \left[\frac{\partial f^l \left(A_a^l \right)}{\partial w_{n_1,n_2}^l} \right] = \frac{\partial}{\partial A_a^l} \left[\frac{\partial f^l \left(A_a^l \right)}{\partial A_a^l} \frac{\partial A_a^l}{\partial w_{n_1,n_2}^l} \right] = \frac{\partial^2 f^l \left(A_a^l \right)}{\partial \left(A_a^l \right)^2} \frac{\partial A_a^l}{\partial w_{n_1,n_2}^l},$$

on rappelle (6.12)

$$\frac{\partial A_a^l}{\partial w_{n_1,n_2}^l} = \begin{cases} X_{n_2}^l & \text{si } a = n_1 \\ 0 & \text{sinon} \end{cases} .$$

Examinons les expressions de la dérivée seconde de la fonction de transfert de la couche l par rapport à l'activation.

• Fonction d'activation sigmoïde On rappelle (B.9)

$$f^{s}(x) = \frac{1}{1 + e^{-x}} \to \frac{\partial f^{s}(x)}{\partial x} = \frac{e^{-x}}{(1 + e^{-x})^{2}},$$

d'où

$$\frac{\partial^2 f^s(x)}{\partial x^2} = \frac{\partial f^s(x)}{\partial x} \cdot \left(\frac{2}{f^s(x)} \cdot \frac{\partial f^s(x)}{\partial x} - 1\right). \tag{6.18}$$

• Fonction d'activation tangente hyperbolique On rappelle (B.10)

$$f^{t}(x) = \tanh(x) \rightarrow \frac{\partial f^{t}(x)}{\partial x} = \left(1 - \tanh^{2}(x)\right),$$

d'où

$$\frac{\partial^2 f^t(x)}{\partial x^2} = -2 \cdot \frac{\tanh(x)}{\cosh^2(x)}.$$
(6.19)

• <u>Fonction d'activation linéaire</u> On rappelle (B.11)

$$f^{l}(x) = x \to \frac{\partial f^{s}(x)}{\partial x} = 1,$$

$$\frac{\partial^{2} f^{l}(x)}{\partial x^{2}} = 0.$$
 (6.20)

d'où

Dans le cas où l'on s'intéresse au jacobien d'une couche l' placée après l dans le réseau, seul le vecteur d'entrée $\vec{X}^{l'}$ est modifié.

$$J^{a,b}\left(T^{l'},\vec{X}^{l'}\right) = w_{a,b}^{l'}\frac{\partial f^{l'}\left(A_a^{l'}\right)}{\partial A_a^{l'}}$$
$$\rightarrow \frac{\partial J^{a,b}\left(T^{l'},\vec{X}^{l'}\right)}{\partial w_{n_1,n_2}^l} = w_{a,b}^{l'}\frac{\partial^2 f^{l'}\left(A_a^{l'}\right)}{\partial w_{n_1,n_2}^l\partial A_a^{l'}} = w_{a,b}^{l'} \cdot \frac{\partial^2 f^{l'}\left(A_a^{l'}\right)}{\partial \left(A_a^{l'}\right)^2} \cdot \frac{\partial A_a^{l'}}{\partial w_{n_1,n_2}^l}$$

La dérivée seconde de la fonction d'activation est la même que calculée précédemment (6.18) (6.19) (6.20). La dérivée de l'activation s'écrit :

$$\frac{\partial A_{a}^{l'}}{\partial w_{n_{1},n_{2}}^{l}} = \sum_{n'=1}^{N_{n'}^{l'}} w_{a,n'}^{l'} \cdot \frac{\partial X_{n'}^{l'}}{\partial w_{n_{1},n_{2}}^{l}} = \sum_{n'=1}^{N_{n'}^{l'}} w_{a,n'}^{l'} \cdot \frac{\partial Y_{n'}^{l'-1}}{\partial w_{n_{1},n_{2}}^{l}}.$$

Cette approche récursive nous demande, tout comme pour l'optimisation sur l'énergie, de connaître la dérivée des vecteurs de sortie de la couche précédente par rapport aux poids : nous l'avons déterminée précédemment (6.14).

En adoptant une démarche analogue, on peut définir la dépendance des forces répulsives par rapport aux biais du réseau. Tout comme précédemment, on a :

$$\frac{\partial F_{j,\alpha}^{\text{rep,NN},i}}{\partial b_{n_1}^l} = -\sum_{k=1}^{N_{\text{atomes}}^i} \frac{\partial}{\partial b_{n_1}^l} \left[J\left(T^{L+1}, \vec{X}^{L+1}\right) \cdot J\left(T^L, \vec{X}^L\right) \cdot \cdots \right. \\ \left. \cdot J\left(T^l, \vec{X}^l\right) \right] \cdot \cdots \cdot J\left(T^0, \vec{G}_{ik}\right) \cdot \frac{\partial G_{ik}}{\partial r_{j,\alpha}^i}.$$

Dans le cas où l'on considère le jacobien de la couche l, on a :

$$\frac{\partial J^{a,b}\left(T^{l},\vec{X}^{l}\right)}{\partial b_{n_{1}}^{l}} = \begin{cases} 0 & \text{si } a \neq n_{1} \\ w_{a,b}^{l} \frac{\partial^{2} f^{l}\left(A_{a}^{l}\right)}{\partial b_{n_{1}}^{l} \partial A_{a}^{l}} = w_{a,b}^{l} \frac{\partial^{2} f^{l}\left(A_{a}^{l}\right)}{\partial \left(A_{a}^{l}\right)^{2}} \cdot \frac{\partial A_{a}^{l}}{\partial b_{n_{1}}^{l}} & \text{si } a = n_{1} \end{cases}$$

Le terme $\frac{\partial^2 f^l(A_a^l)}{\partial (A_a^l)^2}$ est donné plus haut (6.18) (6.19) (6.20). Il nous reste :

$$\frac{\partial A_a^l}{\partial b_{n_1}^l} = \left\{ \begin{array}{ll} 0 & \mbox{ si } a \neq n_1 \\ 1 & \mbox{ si } a = n_1 \end{array} \right. .$$

Pour les couches postérieures l' > l, on a :

$$J^{a,b}\left(T^{l'}, \vec{X}^{l'}\right) = w_{a,b}^{l'} \frac{\partial f^{l'}\left(A_a^{l'}\right)}{\partial A_a^{l'}}$$

$$\rightarrow \frac{\partial J^{a,b}\left(T^{l'}, \vec{X}^{l'}\right)}{\partial b_{n_1}^l} = w_{a,b}^{l'} \frac{\partial^2 f^{l'}\left(A_a^{l'}\right)}{\partial b_{n_1}^l \partial A_a^{l'}} = w_{a,b}^{l'} \cdot \frac{\partial^2 f^{l'}\left(A_a^{l'}\right)}{\partial \left(A_a^{l'}\right)^2} \cdot \frac{\partial A_a^{l'}}{\partial b_{n_1}^l}.$$

Le seul terme à évaluer est :

$$\frac{\partial A_a^{l'}}{\partial b_{n_1}^l} = \sum_{n'=1}^{N_n^{l'}} w_{a,n'}^{l'} \cdot \frac{\partial X_{n'}^{l'}}{\partial b_{n_1}^l} = \sum_{n'=1}^{N_n^{l'}} w_{a,n'}^{l'} \cdot \frac{\partial Y_{n'}^{l'-1}}{\partial b_{n_1}^l}.$$

En somme, cette approche par rétropropagation des gradients depuis la couche de sortie vers la couche d'entrée démontre une praticité certaine à cause de la structure même du réseau, organisé en couches successives. Une complexité supplémentaire émerge néanmoins de l'intrication des nombreuses fonctions et autres opérations performées par les différentes composantes; le recours à la représentation des dérivées par des jacobiens allège le formalisme, et il ne reste qu'à appliquer les théorèmes de dérivations de fonctions composées et de produits. Grâce à tous ces outils, nous avons désormais accès à toutes les grandeurs nécessaires pour optimiser sur l'énergie répulsive (6.4), les forces répulsives (6.5) ou ces deux observables conjointement (6.6).

6.2.2 Filtre de Kalman étendu

Dans cette section, nous présentons dans les grandes lignes l'algorithme du filtre de Kalman étendu, utilisé dans la littérature pour entraîner des réseaux de neurones à reproduire des énergies *DFT*. Cette approche est assez atypique, mais s'avère très intéressante dans des applications similaires à celle de cette thèse où l'on cherche à optimiser des réseaux peu profonds (typiquement quelques couches cachées) avec moins d'une centaine de neurones par couche. Nous nous proposons ici d'en donner une introduction rapide, afin de saisir les spécificités de cette approche.

Originellement [85], l'algorithme du filtre de *Kalman* vise à optimiser un estimateur de système linéaire dynamique : l'idée est de minimiser une fonction coût tenant compte de tous les états du système observés précédemment.

Pour un jeu de données de référence composé de M éléments, on a accès aux $\{\vec{x}_a\}_{a=1}^M$ vecteurs d'entrée, chaque état a étant un état du système dynamique que l'on veut représenter. Pour chaque état \vec{x}_a , on connaît le vecteur de sortie de taille m associé.

$$\vec{y}_a^{\text{ref}} = \left[y_1^{\text{ref}}\left(\vec{x}_a \right), \cdots, y_m^{\text{ref}}\left(\vec{x}_a \right) \right]^T$$

On cherche un estimateur $\vec{y}(t)$, avec N paramètres $\vec{w} = [w_1(t), \cdots, w_n(t)]^T$ tel que $\vec{y}(T) = [y_1(\vec{w}(t), \vec{x}_a), \cdots, y_m(\vec{w}(t), \vec{x}_a)]^T \approx \vec{y}_a^{\text{ref}}$. Pour un état donné du système, il est

donc possible de quantifier la qualité de l'estimateur $\vec{y}(t)$ et de ses paramètres \vec{w} .

$$\vec{\xi}(t) = \vec{y}^{\text{ref}} - \vec{y}(t). \tag{6.21}$$

L'objectif du filtre de Kalman est de minimiser $\Gamma = \sum_{t} \vec{\xi}^{T}(t) \cdot \vec{\xi}(t)$.

Présenté ainsi, l'application à l'optimisation de méthodes d'apprentissage automatique n'est pas évidente : le jeu de données de référence utilisé pour les entraîner n'est pas dynamique, mais fixé pour tout le processus d'optimisation. C'est d'ailleurs l'une des principales raisons justifiant que, assez étonnamment, plusieurs décennies séparent l'avènement du filtre de Kalman de son application en apprentissage automatique.

Pour effectuer ce lien, un léger changement de paradigme est de mise : il faut, à chaque époque, considérer le jeu de données de référence entier comme un système dynamique. Chaque élément du jeu de données de référence est assimilé à un état du système a à un instant t.

Dès lors, on évalue la capacité du réseau à associer un $\vec{y}(t) = E_{\text{rep},a}^{\text{NN}}$ convenable aux vecteurs descripteurs de la boîte $a, \vec{x}_a = \left\{\vec{G}_j\right\}_{j=1}^{N_{\text{atomes}}^a}$. On peut alors réécrire la fonction coût minimisée par l'algorithme comme :

$$\Gamma = \sum_{t=1}^{N_{\text{boites}}} \left(E_{\text{rep},t}^{\text{ref}} - E_{\text{rep},t}^{\text{NN}} \right)^2,$$

qui est une forme assez proche de (6.4), rencontrée plus haut.

Intéressons nous à la manière qu'a l'algorithme du filtre de Kalman d'actualiser les paramètres \vec{w} , dans notre cas les poids et biais du réseau de neurones. À partir de (6.21), on peut obtenir la matrice de dérivées **H**.

$$\mathbf{H}_{ij}(t) = \frac{\partial \xi_j(t)}{\partial w_i}.$$

Cette matrice nécessite de connaître les gradients de la fonction erreur, que l'on peut obtenir facilement avec une démarche très similaire à celle utilisée dans la section précédente (voir. 6.2.1). On peut désormais construire la matrice de gain du filtre de Kalman.

$$\mathbf{K}(t) = \mathbf{P}(t) \cdot \mathbf{H}(t) \cdot \mathbf{A}(t). \tag{6.22}$$

Ici, A est appelée matrice d'échelle, et est définie comme :

$$\mathbf{A}(t) = \left(\frac{1}{\eta(t)}\mathbb{I}_m + \mathbf{H}^T(t) \cdot \mathbf{P}(t) \cdot \mathbf{H}(t)\right)^{-1},$$

où $\eta(t)$ est le taux d'apprentissage, très similaire à celui de l'algorithme de la rétropropagation (6.7), λ , et peut éventuellement évoluer au cours de l'optimisation. Enfin, la matrice $\mathbf{P}(t)$, de dimension $n \times n$ est appelée matrice de covariance de l'erreur, et est ajustée itérativement à chaque état du système.

$$\mathbf{P}(t+1) = \mathbf{P}(t) - \mathbf{K}(t) \cdot \mathbf{H}^{T}(t) \cdot \mathbf{P}(t) + \mathbf{Q}(t),$$

que l'on initialise comme $\mathbf{P}(0) = \epsilon^{-1} \mathbb{I}_n$ à partir d'un paramètre ϵ fixé par l'utilisateur. Enfin, $\mathbf{Q}(t)$ est une matrice faisant office de bruit d'échantillonnage, permettant d'éviter le sur-apprentissage et de sortir plus facilement des minima locaux.

$$\mathbf{Q}(t) = q(t)\mathbb{I}_n.$$

Après définition des termes intermédiaires, nous pouvons utiliser la matrice de gain (6.22) pour actualiser les paramètres du modèle :

$$\vec{w}(t+1) = \vec{w}(t) + \mathbf{K}(t) \cdot \xi(t)$$

Ainsi, on actualise tous les paramètres M fois (pour $\vec{w}(t), t = [0, M]$) en parcourant l'intégralité du jeu de données de référence : cela constitue **une** époque d'apprentissage. Lors du passage à l'époque suivante, on repart de t = 0 et on réitère la processus. Autrement dit, t = a désigne ici l'élément du jeu de données de référence que l'on considère, et non l'époque d'apprentissage dans laquelle on se trouve.

Le processus d'apprentissage d'un réseau de neurones en utilisant l'algorithme du filtre de Kalman est en réalité plus nuancé, et de nombreuses subtilités interviennent, notamment lorsque l'on souhaite inclure les forces dans la fonction coût. Nous nous contentons ici de la présentation ci-dessus, permettant de saisir le fonctionnement global de l'approche. Plus de détails sont disponibles dans la littérature, notamment dans la présentation des caractéristiques d'apprentissage du logiciel N2P2 [86], utilisé lors des travaux de cette thèse.

Chapitre 7

Construction d'un potentiel répulsif numérique

Lors des chapitres précédents, nous avons présenté l'intégralité des outils et concepts permettant de comprendre les différents détails et enjeux des présents travaux. Ce chapitre s'appuie sur ceux-ci, et aspire à rendre compte du processus permettant de générer un potentiel répulsif à même d'être utilisé au sein d'un code DFTB. Pour cela, nous nous intéressons à la construction d'un potentiel répulsif global pour le Silicium pur. Dans un premier temps, nous portons notre attention sur l'obtention des données de référence employées pour entraîner les réseaux de neurones (voir 7.1) : quels phases thermodynamiques choisir ? Quels calculs effectuer ? Quel traitement appliquer aux données ? Ensuite, nous présentons le processus d'apprentissage en lui-même (voir 7.2), en discutant notamment de l'évaluation *a priori* de sa qualité et de quelques problèmes qui lui sont propres.

7.1 Génération des données de référence

La flexibilité des méthodes d'apprentissage automatique occulte parfois une vérité simple : le facteur déterminant le plus la qualité de prédiction est, au final, le jeu de données de référence. Ces outils excellent en interpolation, mais ne sont que très rarement performants en extrapolation¹. En conséquence, l'adage selon lequel on n'est «jamais meilleur que la base d'apprentissage» est plus que justifié. Un effort maximal doit être fourni pour la génération de données de référence, et les travaux effectués dans le cadre cette thèse ne font pas exception.

7.1.1 Choix des systèmes

Dans l'optique d'éprouver les capacités d'apprentissage et de prédiction de la méthodologie présentée plus haut, nous avons choisi de nous intéresser au comportement du

^{1.} Euphémisme par courtoisie envers l'existence de l'apprentissage par transfert.

Silicium pur.

Ce choix nous semble pertinent pour plusieurs raisons. D'une part, il s'agit évidemment d'un matériau très utilisé dans pléthore d'applications. D'autre part, cet élément est bien connu de la communauté s'intéressant à la simulation numérique, et une quantité solide d'informations est à notre disposition pour comparer et évaluer nos potentiels. Enfin, cela nous permet d'explorer un diagramme de phase complexe, avec de nombreuses phases stables ou métastables mais aussi des nano-structures.

Notons pour commencer un point capital : à l'inverse des potentiels de paire, nous ne cherchons pas à reproduire l'énergie de liaison de systèmes connus mais des énergies répulsives totales de boîte. Dès lors, on peut se permettre d'ajouter des systèmes très variés aux données utilisées pour optimiser les paramètres du potentiel. Les potentiels de paire se restreignent soit à des systèmes isolés soit très symétriques, de manière à identifier clairement l'énergie répulsive d'**une** paire d'atomes. Ce procédé est fastidieux, requiert une grande attention de l'utilisateur et, bien qu'il soit en principe possible d'optimiser un potentiel de paire pour n'importe quel système de référence, s'avère bien trop difficile pour permettre ceci en pratique pour des systèmes complexes. Notre approche se libère complètement de cette contrainte, et on peut en théorie s'intéresser à n'importe quel système sans discrimination, fut-il périodique ou non, désordonné ou cristallin.

Le Silicium présente plusieurs phases condensées, et nous choisissons d'en intégrer un maximum dans notre jeu de données de référence. Tout d'abord, on identifie des polymorphes cristallins de symétries différentes : diamant, hexagonale, cubique (Weaire-Phelan), monoclinique et β -Sn. Par ailleurs, l'inclusion de phases désorganisées est souhaitable, sous la forme de liquides et d'amorphes.

Nous souhaitons également nous intéresser spécifiquement à un problème bien connu des utilisateurs de méthodes d'apprentissage automatique en simulation atomistique. Il est notoirement difficile de représenter avec un même outil le comportement des systèmes périodiques (tels que les cristaux) et isolés (tels que les molécules ou les nano-structures). L'étude de ces derniers avec des outils d'apprentissage automatique requiert une description spécifique, axée sur la géométrie globale des molécules et sur l'agencement des atomes. Cette approche est très pertinente dans le cas de systèmes isolés car le nombre d'atomes est limité et les états quantiques sont localisés sur les atomes, et peuvent induire des charges ponctuelles perçues aux longues distances : un mouvement sensible d'un groupe d'atomes d'un côté de la molécule affecte alors la configuration de celle-ci manière non négligeable, même à longue portée. À l'inverse, la considération de systèmes périodiques (*infinis*) impossibilise une approche à la granularité aussi fine : pour simuler un système (par exemple, du Silicium diamant) l'on parvient souvent à rendre compte des propriétés du système à partir de contributions localisées. On utilise dans ce cas les potentiels numériques «classiques» [75,84], qui sont construits pour reproduire les énergies de systèmes périodiques, en se limitant à des contributions courte portée pour construire des descripteurs adaptés à des environnements très variables.

Notre approche est différente : on s'intéresse uniquement aux interactions répulsives de très coutre portée, et les interactions électrostatiques à longue portée sont représentées par le terme de second ordre (3.9) du formalisme DFTB, dont la robustesse théorique

permet de considérer une large variété de systèmes [24], périodiques [87] ou non [88,89]. On intègre donc au jeu de données de référence plusieurs objets isolés, afin de les représenter conjointement aux phases périodiques. Premièrement, on inclut des **dimères** Si₂ pour lesquels on fait varier la distance interatomique : il s'agit historiquement (Chap. 3) des systèmes utilisés pour générer des potentiels répulsifs. Nous avons également à disposition [90] un jeu de **nano-clusters** aux configurations variées dont la taille est comprise entre 14 et 77 atomes : via leur inclusion, on espère représenter précisément ces objets et leurs analogues.

7.1.2 Calculs de structure électronique

L'objectif est de générer un jeu de données de référence liant les positions des atomes dans une boîte de simulation (converties en environnements atomiques \mathbf{R}_i) aux observables nous intéressant : énergie répulsive de boîte $E_{\rm rep}^{\rm ref}$ et forces répulsives agissant sur ces atomes $F_{i,\alpha}^{\rm rep,ref}$.

On cherche à entraîner un réseau de neurones à représenter le jeu de données de référence. Pour chacune des boîtes que ce dernier contient, on veut calculer :

$$E_{\rm rep}^{\rm tot,NN} = \sum_{i=1}^{N_{\rm atomes}} E_{{\rm rep},i}^{\rm NN} \approx E_{\rm rep}^{\rm tot,ref} \qquad \text{et} \qquad F_i^{\rm rep,NN} \approx F_i^{\rm rep,ref}.$$
 (7.1)

En conséquence, il nous faut avoir accès aux références $E_{\rm rep}^{\rm tot,ref}$ et $F_i^{\rm rep,ref}$. On rappelle que (6.2) :

$$E_{\rm rep,boite}^{\rm ref} = E_{\rm DFT}^{\rm ref} - E_{\rm DFTB, E_{\rm reg}}^{\rm ref},$$

et on définit les forces répulsives de référence comme suit.

$$F_{i,\alpha}^{\text{rep,ref}} = -\frac{\partial E_{\text{rep,boite}}^{\text{ref}}}{\partial r_{i,\alpha}} = -\frac{\partial E_{\text{DFT}}^{\text{ref}}}{\partial r_{i,\alpha}} + \frac{\partial E_{\text{DFTB},\overline{E}_{\text{reg}}}^{\text{ref}}}{\partial r_{i,\alpha}} = F_{i,\alpha}^{\text{DFT,ref}} - F_{i,\alpha}^{\text{DFTB,ref},\overline{E}_{\text{reg}}}.$$
 (7.2)

A partir de là, deux étapes se dessinent : tout d'abord, il nous faut obtenir plusieurs boîtes de simulation représentatives des systèmes que l'on veut couvrir. On cherche à connaître, pour chacune d'elles, les positions atomiques, l'énergie totale DFT de la boîte $E_{\text{DFT}}^{\text{ref}}$ et les forces atomiques DFT associées $F_{i,\alpha}^{\text{DFT,ref}}$.

Par la suite, un calcul DFTB doit être effectué pour chacune de ces boîtes, afin d'obtenir l'énergie totale DFTB sans potentiel répulsif $E_{\text{DFTB}, \overline{E_{\text{reg}}}}^{\text{ref}}$ et les forces DFTBsans potentiel répulsif $F_{i,\alpha}^{\text{DFTB,ref}, \overline{E_{\text{reg}}}}$.

7.1.2.1 Calculs VASP

Nous utilisons le logiciel **VASP** [91–93] pour nos calculs de structure électronique DFT. En amont de la génération des données de références, mettons en avant quelques points clés qui définissent les calculs à effectuer.

D'une part, on souhaite que le jeu de données de référence soit le plus divers possible et qu'il inclue donc une grande variabilité configurationnelle et énergétique. Il n'est d'ailleurs pas nécessaire, et serait même préjudiciable, d'avoir des données trop ressemblantes en son sein : éviter de telles redondances équivaut à s'assurer de considérer des boîtes de simulation décorrélées. D'autre part, on espère pouvoir représenter une large variété de systèmes et couvrir une majorité des applications physiques pour auxquelles nos outils seront employés. Cette assertion comprend des systèmes hybrides ou transitoires ; il est donc capital que **tous** les éléments du jeu de données de référence soient consistants : c'est-à-dire générés avec un maximum de paramètres similaires, d'approximations identiques. Dans le cas d'un calcul **VASP**, cela se traduit par :

- l'utilisation de la même fonctionnelle d'échange-corrélation. Dans notre cas, on choisit l'approximation PBE (voir 2.3.2), pour sa robustesse, son coût relativement réduit et sa fiabilité pour une large gamme de systèmes. La méthode PAW (Projector Augmented Wave PAW_PBE Si 05Jan2001) [94,95] est employée pour son bon compromis entre coût numérique et qualité de description de la fonction d'onde aux très courtes distances, capitales pour notre application par rapport aux autres pseudo-potentiels et aux méthodes tous-électrons.
- l'emploi du même jeu de fonctions de base. Le logiciel utilisant des ondes planes, nous observons une précision satisfaisante pour ENCUT = 400 eV, qui spécifie l'énergie maximale du jeu de bases d'ondes planes. À titre indicatif, l'impact des paramètres ENCUT = 400 eV ou ENCUT = 520 eV sur l'énergie minimale des monocristaux n'excède jamais 0.3 meV/atome, mais fait varier d'environ un facteur deux le temps de calcul. Cette limite nous semble donc être un compromis raisonnable entre robustesse et temps de calcul.
- la constance de la prise en compte de la polarisation. Pour les systèmes nous intéressant, nous avons observé que la polarisation induite par les spins du systèmes a un impact marginal, et on choisit de s'en passer (ISPIN = 1) afin de réduire le temps de calcul. Pour information, l'impact de la prise en compte de la polarisation du spin sur l'énergie minimale des monocristaux n'excède pas 0.9 µeV/atome, mais change approximativement les temps de calcul du simple au double.
- le même seuil de convergence des calculs auto-cohérents à EDIFF = 0.1 meV; ce choix, non optimal pour certains systèmes très stables, est capital afin d'assurer une convergence robuste pour tous les cas envisageables.

En un sens, ces choix fixent le seuil de précision maximal pour notre potentiel répulsif par réseau de neurones, et semblent sous-cet angle acceptables. On est alors en capacité d'effectuer des calculs dans l'intention d'extraire pour chaque boîte de simulation :

$$\left\{\mathbf{R}_{i}, \ \vec{F}_{i}^{\mathrm{DFT,ref}}\right\}_{i=1}^{N_{\mathrm{atomes}}}, \quad E_{\mathrm{DFT}}^{\mathrm{ref}}.$$

Lorsque l'on aspire à générer un potentiel numérique, on effectue habituellement un calcul de dynamique moléculaire pour chaque système d'intérêt. Ceux-ci s'effectuent à une température élevée, sans être excessive pour ne pas représenter que des liquides. Ce choix de température s'explique facilement en considérant la probabilité p_i d'observer un état quelconque *i* du système, d'énergie ε_i . Cette probabilité suit une distribution de Boltzmann [96], et dépend donc explicitement de la température :

$$p_i \propto e^{\frac{-\varepsilon_i}{k_B T}},\tag{7.3}$$

où k_B est une valeur constante appelée constante de Boltzmann. D'après (7.3), aux basses températures les configurations très stables (petit ε_i) sont les seules observées; on augmente la probabilité de voir apparaître des configurations moins stables énergétiquement en chauffant le système.

Cela s'intuitionne assez facilement : plus la température augmente, plus l'énergie cinétique des atomes croît et plus ils sont à même de sortir de leurs minima d'énergie. Dès lors, on échantillonne en principe une variété bien plus grande de configurations en se plaçant à plus haute température. Si la température augmente néanmoins trop, la distribution de probabilités p_i est dominée par la température et l'impact de l'énergie ε_i s'amenuise; la distribution (7.3) s'uniformise alors, et la fréquence d'apparition des états stables diminue d'autant - ceux-ci sont alors noyés au milieu de configurations plus désordonnées.

Quand le système est un solide - dans notre cas, les cristaux ou les amorphes - on se place usuellement légèrement en dessous de sa température de fusion. On reste représentatif de l'état fondamental du système; les atomes oscillent en moyenne autour de leur position d'équilibre, et des environnements peu perturbés sont également présents dans les données générées. Voici, le détail des dynamiques effectuées :

- pour les symétries β -Sn (64 atomes), cubique (Weaire-Phelan 46 atomes), hexagonale (64 atomes) et monoclinique (96 atomes), on effectue une dynamique sur 1 ps à T = 1000 K et deux dynamiques à T = 300 K sur 1 ps, dont une à haute pression.
- pour la symétrie diamant, le calcul s'effectue sur 64 atomes en deux temps. Tout d'abord, une première dynamique sur 0.5 ps à T = 300 K, puis une seconde sur 0.1 ps à T = 2000 K afin de beaucoup perturber les environnements observés et d'observer l'apparition locale de symétries légèrement désordonnées.
- Des dynamiques relativement courtes (respectivement 0.2 ps, 0.25 ps et 0.5 ps) à T = 2000 K pour les amorphes (93, 96, 100 atomes). Le processus suivi pour obtenir les configurations amorphes initiales est présenté en détail plus loin (voir 8.3.2).

Dans le cas des liquides, cet embarras disparaît et on choisit de se placer à très haute température², T = 3000 K sur 6 ps pour 128 atomes afin d'explorer longuement l'espace des configurations.

^{2.} Sans pour autant excéder la température d'ébullition du Silicium pur, T = 3565 K.

Conjointement aux calculs de dynamique moléculaire, on choisit d'employer une seconde approche, en faisant évoluer graduellement la distance entre les atomes. Dans le cas du dimère, le processus est trivial : on place initialement les deux atomes de silicium à une très courte distance, dans notre cas r = 0.8 Å, et on augmente graduellement la distance jusqu'à atteindre $r = r_{cut}$. Cette démarche s'adapte également aux cristaux : il est ici question d'obtenir des courbes énergies vs. volume. Le calcul s'effectue en mettant à l'échelle une boîte de simulation d'un cristal à l'équilibre par un facteur compris entre 0.85 et 1.15. On multiplie donc les vecteurs de boîte, les positions, et en conséquence les distances, par ce facteur d'échelle. En faisant cela, on inclut également des données de référence à des pressions très variées.

Enfin, l'étude [90] propose pour chaque taille de nano-cluster plusieurs configurations variées (dont le processus d'obtention est détaillé plus loin - voir 8.2.1). Ce dispositif s'intègre parfaitement à notre méthodologie : cela garantit d'allier diversité des systèmes et stabilité physique. On inclut donc ces clusters dans notre jeu de données de référence, en optimisant au préalable leur géométrie avec VASP. Afin d'explorer partiellement le paysage énergétique autour de ces configurations, il est également souhaitable de les distordre légèrement. Les nano-objets peuvent être instables si la température augmente ; on préfère donc à la dynamique moléculaire une approche par compression/extension où l'on met là encore à l'échelle les distances interatomiques.

7.1.2.2 Calculs DFTB+

Il nous faut également déterminer pour chacune des boîtes de simulation $\{\mathbf{R}_i\}_{i=1}^{N_{\text{atomes}}}$ l'énergie *DFTB* sans interaction répulsive et les forces associées.

$$\left\{ \mathbf{R}_{i}, \ \vec{F}_{i}^{\mathrm{DFTB, ref}, \mathcal{E}_{\mathrm{reg}}} \right\}_{i=1}^{N_{\mathrm{atomes}}} , \quad E_{\mathrm{DFTB}, \mathcal{E}_{\mathrm{reg}}}^{\mathrm{ref}}.$$

Ici, le processus à suivre est plus simple. On modifie dans un premier temps les fichiers de *Slater-Koster* utilisés de manière à nullifier toute interaction répulsive. Dès lors, il nous suffit, pour chaque boîte, d'effectuer un calcul statique unique et de récolter les grandeurs désirées : nous sommes alors en capacité de calculer les différences (6.2) et (7.2) pour chaque configuration.

Les paramètres des calculs étudiés sont là aussi consistants :

- On utilise les paramétrisations des termes électroniques de l'énergie *DFTB* pour le Silicium issues de la thèse d'Alexander Sieck [36].
- Le seuil de convergence des charges lors des calculs auto-cohérents est volontairement très faible (SCCTolerance = 1×10^{-8} e, où e est la charge de l'électron) de manière à assurer une convergence - et donc une précision - optimale. Le temps de calcul *DFTB* étant assez faible de manière générale, cette rigueur supplémentaire n'impacte pas le temps de calcul de manière trop dommageable.
- On fixe le moment angulaire maximal (MaxAngularMomentum = p).

7.1.3 Translation des énergies répulsives

En pratique, une nuance est tue dans notre présentation de l'équation (3.15). Dans l'absolu, DFT et DFTB n'expriment pas leurs énergies respectives sur la même échelle : seules leurs variations relatives d'énergie le sont.

Au sein des calculs VASP, les énergies reportées sont essentiellement des énergies de liaison; pour cela, le code soustrait en interne les énergies individuelles de chaque atome. La valeur de ces énergies individuelles, $E_{\rm at}^{\rm PP}$, est spécifiée dans les paramètres du pseudopotentiel utilisé.

$$E^{\text{VASP}} = E_{\text{tot}} - N \times E_{\text{at}}^{\text{PP}}$$

Cette définition vise en principe à imposer que l'énergie d'un atome isolé soit nulle : dans les faits, on parvient à la rendre très faible, mais pas exactement nulle, $E_{\rm at}^{\rm VASP} \neq 0$.

L'approche semi-empirique de DFTB est différente : d'une part, le formalisme ne tient formellement pas compte des électrons de cœur dans le calcul de l'énergie totale. Par ailleurs, l'énergie de boîte retournée est l'énergie totale. En conséquence, cette énergie est très différente de zéro, même pour un atome isolé. On remarque ainsi que les énergies retournées par les deux méthodes n'évoluent pas par rapport à la même référence, et n'ont donc pas le même zéro d'énergie.

Dès lors, le potentiel répulsif de référence que l'on calcule comme leur différence est très grand, et reste de l'ordre de dizaines d'eV même dans le cas d'un unique atome n'interagissant pas. L'idée est de translater les énergies *VASP* et *DFTB* de manière leur attribuer la même échelle d'énergie.

Notons que ce choix est entièrement arbitraire; il est surtout motivé par la définition même des interactions répulsives, voulant que ces dernières s'annulent à mesure que les atomes s'éloignent, et l'on choisit donc l'atome isolé comme zéro d'énergie. Il est donc requis de déduire les énergies d'atomes isolés 3 VASP ($E_{\rm at}^{\rm VASP}$) et DFTB ($E_{\rm at}^{\rm DFTB}$). Le potentiel répulsif, après cet ajustement, devient donc :

$$E_{\rm rep}^{\rm ref, a just.} = E_{\rm rep}^{\rm ref} - N \left(E_{\rm at}^{\rm VASP} - E_{\rm at}^{\rm DFTB} \right).$$

Remarque : il nous faut effectuer cette translation à la main, car rien dans notre construction méthodologique ne contraint nos outils à associer des vecteurs de sortie nuls à un atome isolé. Ainsi, et même si le vecteur descripteur d'un tel environnement sera forcément nul, les biais de chaque neurones du réseau n'ont aucune raison particulière de valoir zéro : on peut (et c'est sans doute souvent le cas en pratique) attribuer une valeur de sorte non nulle à un vecteur descripteur ne contenant que des zéros.

Cette modification semble conceptuellement satisfaisante; dans le cas d'atomes réellement isolés $(r \to \infty)$, cette translation assure en théorie une énergie répulsive nulle. En

^{3.} Celles-ci sont calculées pour un système non périodique avec *DFTB*, et avec une boîte de très grande taille dans *VASP*.

pratique les énergies répulsives en résultant sont raisonnables et proches de zéro, mais cet ajustement est perfectible : pour un dimère avec une distance interatomique d'environ $r \approx r_{\rm cut}$, l'énergie $E_{\rm rep}^{\rm ref, ajust.}$ est légèrement négative.

$$E_{\text{rep,dimère}}^{\text{ref,ajust.}}\left(r_{\text{cut}}\right) = O\left(-10^{-1}\right) \text{ eV} < 0.$$

Ici, $r_{\rm cut}$ réfère au rayon de coupure des fonctions de symétrie (voir 5.4.1). Le fait que l'énergie translatée soit non-nulle est intrinsèquement liée à notre rayon de coupure fini; dans la limite $r_{\rm cut} \to \infty$, l'équation précédente serait vérifiée. Il n'est néanmoins pas souhaitable d'augmenter à volonté le rayon de coupure : l'énergie répulsive est un terme hautement localisé, et les différences d'énergie observées par rapport à VASP pour les atomes plus éloignés sont dues aux imprécisions des contributions électroniques à l'énergie *DFTB*. En conséquence, une valeur trop grande du rayon de coupure entrave la transférabilité du potentiel répulsif résultant.

Dans notre formalisme, les interactions au-delà de $r_{\rm cut}$ n'ont aucune importance. Un atome sans aucun voisin ($r \ge r_{\rm cut}$) est ainsi considéré comme isolé du point de vue de l'énergie répulsive. Un choix de zéro d'énergie plus adéquat est donc de s'assurer que l'énergie répulsive de référence dans le cas où $r = r_{\rm cut}$ soit rigoureusement nulle. On peut par exemple translater l'énergie de référence proportionnellement à l'énergie d'un dimère pour lequel $r = r_{\rm cut}$:

$$E_{\rm rep}^{\rm ref,final} = E_{\rm rep}^{\rm ref,ajust.} - \frac{N}{2} E_{\rm rep,dimer}^{\rm ref,ajust.} (r_{\rm cut}), \qquad (7.4)$$

que l'on écrit par la suite $E_{\rm rep}^{\rm ref}$ par commo dité.



FIGURE 7.1 – Schématisation du processus de translation des énergies répulsives de référence.

Notons au passage que cette construction pourrait s'appliquer en prenant en compte d'autres systèmes qu'un simple dimère ; par exemple, on pourrait tout aussi bien translater de $E_{\text{rep,diamant}}^{\text{ref,ajust.}}(r_{\text{cut}})$, dans le cas d'une symétrie diamant où tous les voisins sont distants de $r = r_{cut}$. Il paraît toutefois peu probable de rencontrer *in-situ* de telles configurations, et leur description exacte semble peu importante. À l'inverse, on imagine sans trop de peine le cas où un atome n'a qu'un seul voisin distant de r_{cut} au sein d'un système isolé ; dans de telles situations, la translation (7.4) est souhaitable. On peut trouver une représentation schématique de ce processus de sélection dans la figure 7.1.

Soulignons qu'il aurait été identique de translater en une fois les énergies de manière à atteindre $E_{\text{rep,dimère}}^{\text{ref}}(r_{\text{cut}}) = 0$. Néanmoins, la procédure présentée nous permet clairement d'identifier la part de cette translation que l'on doit à la construction sémantique des énergies au sein de *VASP* et *DFTB*, qui est justifiée physiquement, et celle plus artificielle qui *«compense»* des imprécisions numériques dues aux contributions électroniques de *DFTB* et au rayon de coupure fini.

Remarquons finalement l'indépendance des forces répulsives par rapport à cette translation en énergie répulsive. Ajouter ou retirer une constante à une fonction ne modifie pas ses dérivées ; aucune translation des forces n'est requise.

7.1.4 Sélection des données de référence

Nous avons désormais un large jeu de référence $\Omega^{\rm ref}$ à notre disposition, qui contient pour chaque boîte :

$$\left\{ \mathbf{R}_{i}, \ \vec{F}_{i}^{\mathrm{rep,ref}} \right\}_{i=1}^{N_{\mathrm{atomes}}}, \quad E_{\mathrm{rep}}^{\mathrm{ref}}$$

Nous choisissons de n'en utiliser qu'une partie Ω^{sel} pour entraîner nos outils. D'une part, deux configurations successives d'une dynamique moléculaire sont très similaires, et un excès de redondances dans les données de référence peut masquer des phénomènes plus rares. D'autre part, on devine qu'un jeu de données volumineux alourdit fortement le coût numérique du processus d'apprentissage.

Usuellement, la sélection des configurations est aléatoire, et on s'assure uniquement de leurs faibles corrélations respectives. Cette approche est pertinente et fonctionne de manière satisfaisante. Il est néanmoins légitime d'imaginer un choix plus efficace, par exemple quand la base de données contient plusieurs variations d'un même système. Pour cela, on se concentre sur les deux grandeurs principales déterminantes dans l'importance d'une boîte.

D'une part, l'énergie répulsive de référence est centrale : c'est, *in fine*, la grandeur que l'on veut reproduire et il semble logique de vouloir assurer une variété maximale des valeurs cibles. Évidemment, il convient de tenir compte du nombre d'atomes dans la boîte; on sélectionne pour commencer les boîtes d'énergies de référence par atome extrémales.

$$\Omega^{sel} = \Omega_m^{\text{ref}} + \Omega_M^{\text{ref}} \quad , \qquad m = \underset{i \in \Omega^{\text{ref}}}{\operatorname{argmin}} \frac{E_{\text{rep}}^{\text{ref},i}}{N_{\text{atomes}}^i} \quad \text{et} \qquad M = \underset{i \in \Omega^{\text{ref}}}{\operatorname{argmax}} \frac{E_{\text{rep}}^{\text{ref},i}}{N_{\text{atomes}}^i}.$$

Ensuite, on identifie itérativement les boîtes aux énergies les plus distantes de celles déjà retenues.

$$\Omega^{\text{sel}} \to \Omega^{\text{sel}} + \Omega_s^{\text{ref}} , \qquad s = \underset{\substack{i \in \Omega^{\text{ref}} \\ i \notin \Omega^{\text{sel}}}}{\operatorname{argmax}} \sqrt{\sum_{j \in \Omega^{\text{sel}}} \left(\frac{E_{\text{rep}}^{\text{ref},i}}{N_{\text{atomes}}^i} - \frac{E_{\text{rep}}^{\text{ref},j}}{N_{\text{atomes}}^j}\right)^2}.$$

On arrête la procédure quand on atteint un nombre défini de boîtes, ou bien quand les distances en énergies par atome deviennent assez petites.

On décide aussi de s'intéresser à la distribution des distances : il s'agit de la grandeur clé dans le calcul des fonctions de symétrie, et donc des potentiels répulsifs. Comme précédemment, l'objectif est d'assurer une diversité maximale de distances au sein du jeu de données de référence.

Un bon outil pour jauger de la diversité des distances présentes au sein d'**une** boîte est la **fonction de distribution radiale** g(r), qui est proportionnelle à la probabilité de trouver un voisin à une distance donnée d'un atome. En pratique, il s'agit d'un histogramme (Fig. 7.2) recensant les distances du systèmes par tranches de distances d'épaisseur fixe.

À tout instant d'un processus de sélection, on peut additionner les dis-



FIGURE 7.2 – Exemple de fonction de distribution radiale, construite comme un histogramme.

tributions g(r) de **toutes** les boîtes sélectionnées, et obtenir un **histogramme cumulé**. On peut évaluer sa **platitude** (*flatness*) afin de quantifier la diversité globale des distances des systèmes déjà sélectionnés. Plus un histogramme est plat, et plus les valeurs qu'il contient sont distribuées uniformément. Dès lors, un histogramme très plat correspond dans notre cas à une bonne diversité de distances.

On calcule la platitude de la fonction de distribution radiale cumulée grâce à la méthode χ^2 [97]. Il s'agit de la mesure de l'écart d'un histogramme à sa valeur moyenne tpour chacune de ses N intervalles.

$$\chi^2 = \sqrt{\sum_{i=1}^{N} \frac{(h_i - t)^2}{t}}, \ t = \frac{1}{N} \sum_{i=1}^{N} h_i.$$

Ici, h_i est la valeur du $i^{\text{ème}}$ intervalle de l'histogramme.

Une fois en mesure d'évaluer à tout instant la diversité des distances cumulées de plusieurs boîtes, on est en mesure de déterminer parmi plusieurs candidates **la** boîte dont l'ajout aux données de référence sera le plus bénéfique. En pratique, on sélectionne en première la boîte dont le score individuel χ^2 est le plus faible parmi toutes les boîtes candidates. À partir de là, on peut avoir une approche itérative :

- calcul pour chaque boîte candidate restante le nouveau score χ^2 hypothétique si on l'ajoute aux données sélectionnées.
- si c'est possible, sélection de la boîte qui fait le plus diminuer χ^2 , et on repart au premier point.
- si c'est impossible (tout nouvel ajout ne ferait qu'augmenter χ^2) ou que l'on a sélectionné suffisamment de boîtes (fixé par l'utilisateur), on arrête le processus.

Il existe des distances préférentielles clairement marquées pour un système donné et l'immense majorité des boîtes ont une fonction de distribution radiale relativement similaire aux autres. En conséquence, cette sélection par distribution de distances est limitée, mais elle est notamment intéressante pour présélectionner les systèmes en assurant une proportion significative de courtes distances. Celles-ci sont rares dans le jeu de données de référence Ω^{ref} mais très précieuses pour notre application. En effet, ce sont elles qui permettent aux réseaux de neurones de construire des potentiels répulsifs pénalisant les rapprochements excessifs d'atomes.

Les deux approches présentées (sélection de boîtes par énergie ou par distances) sont assez complémentaires. En pratique, il est intéressant de présélectionner dans un premier temps en fonction de la distance : on ne récupère que quelques boîtes (de l'ordre de la dizaine), mais celles-ci assurent de contenir quelques distances interatomiques réduites. Dans un second temps, on performe une sélection par énergie au sein des boîtes restantes de manière à finir de remplir le jeu de données de référence.

7.1.5 Le cas des très faibles distances

Indépendamment des données de référence choisies et du processus de sélection opéré, certains domaines de la surface d'énergie potentielle ne peuvent pas être inclus pour l'apprentissage. Le cas des très courtes distances en fait notamment partie : même en tentant d'inclure des courbes énergie vs volume ou des dimères où l'on rapproche les atomes, il existe une limite au-delà de laquelle nous ne pouvons pas obtenir de données. Par exemple, nous incluons pour la paire d'atomes Si_2 des distances d'au moins 0.8 Å. En deçà de cette distance, il est ardu d'effectuer des calculs; on pourrait d'ailleurs discuter de la pertinence même de tels calculs, pour lesquels orbitales atomiques se superposent et qui n'apparaissent *in-situ* que dans des circonstances inhabituelles. Plus vulgairement,

ces situations n'arrivent en pratique que si quelque chose au sein de la simulation s'est «mal passé».

Ce manque de données de référence n'est pas problématique dans le cas des potentiels répulsifs de paire, pour lesquels on peut définir une forme purement convexe, strictement croissante aux faibles distances pour prohiber les chevauchements entre atomes. Lorsque l'on emploie des outils d'apprentissage automatique, il est impossible de prédire *a priori* le comportement qu'ils adopteront dans le cas de systèmes pour lesquels aucune référence n'est disponible. Dans le même temps, on comprend bien comment il serait problématique de permettre un rapprochement excessif d'atomes : même si ces situations sont dans les faits très irrégulières, il est important que le potentiel répulsif assure effectivement la répulsion en pénalisant ces occurrences.

L'exactitude de la répulsion semble néanmoins assez secondaire; on propose dans ces travaux de s'orienter vers une pénalisation exponentielle si la distance devient inférieure à une valeur seuil r_0 .

$$E_{\rm rep}(r) = \begin{cases} E_{\rm rep}^{\rm NN}(r) & \text{si } r \ge r_0, \\ \\ E_{\rm rep}^{\rm short}(r) &= a \cdot e^{-r} + b \text{ sinon.} \end{cases}$$
(7.5)

Pour fixer les valeurs de a et b, on choisit d'assurer une continuité des énergies et forces pour un dimère avec $r = r_0$. Dans le cas de deux atomes tels que $r_{\text{atome}_1,y} = r_{\text{atome}_2,y}$ et $r_{\text{atome}_1,z} = r_{\text{atome}_2,z}$ - c'est-à-dire que le déplacement ne se fait que selon la direction x on cherche à assurer que $E_{\text{rep}}^{\text{short}}(r_0) = E_{\text{rep}}^{\text{NN}}(r_0)$ et $F_{\text{rep},x}^{\text{short,dimère}}(r_0) = F_{\text{rep},x}^{\text{NN}}(r_0)$. On a donc

$$a = F_{\text{rep},x}^{\text{NN}}(r_0) \cdot e^{-r_0} \cdot \frac{r_{x,0}}{r_0}$$

et $b = E_{\text{rep}}^{\text{NN}}(r_0) - a \cdot e^{-r_0},$

avec $r_{x,0} = r_{\text{atome}_{1,x}} - r_{\text{atome}_{2,x}}$ quand $r = r_0$. La détermination de $E_{\text{rep}}^{\text{short}}$ pour un dimère est alors aisée. Les composantes $\alpha = \{x, y, z\}$ des forces interatomiques s'expriment pour leur part

$$F_{\text{rep},i,\alpha}(r) = \begin{cases} F_{\text{rep},i,\alpha}^{\text{NN}}(r) & \text{si } r \ge r_0, \\ \\ F_{\text{rep},i,\alpha}^{\text{short}}(r) &= a \cdot e^{-r} \cdot \frac{\partial r}{\partial r_{i,\alpha}} & \text{sinon.} \end{cases}$$
(7.6)

Dans le cas de systèmes autres que les dimères, où les atomes ont plusieurs voisins, l'idée est de traiter les éventuels cas où des atomes sont plus proches que r_0 en retirant le(s) voisin(s) trop proche(s) et en le(s) traitant à part, comme présenté dans la figure 7.3.

Cette approche n'est manifestement plus aussi adaptée quand plusieurs voisins entrent en compte que dans le cas d'un dimère : notre potentiel répulsif est, par construction, à plusieurs corps, et ne peut pas être réduit à une somme de contributions à deux corps.



FIGURE 7.3 – Schéma du fonctionnement du potentiel répulsif ad-hoc de courte portée.

On introduit donc inévitablement une discontinuité, dont l'importance est néanmoins à relativiser. Ce potentiel *ad-hoc* a courte portée n'est pas supposé être requis au cours de simulations *normales*; il ne s'agit **que** d'un filet de sécurité, d'un dernier recours n'intervenant que dans des situations très irrégulières, où l'on considère des systèmes n'ayant de toute façon que peu de sens physique. Il est nécessaire pour cela de s'assurer que r_0 soit petit, idéalement $r_0 < 2r_{\text{atome}}$, où r_{atome} est le rayon atomique de l'espèce considérée. En raison de la possible discontinuité qu'elle apporte, cette correction aux très courtes portées est optionnelle (voir annexe C.2).

7.2 Apprentissages

Intéressons nous désormais aux processus d'apprentissage effectués. Si l'objectif est, au final, d'obtenir un potentiel répulsif **général** applicable à un maximum de systèmes, plusieurs étapes intermédiaires sont détaillées : en partant d'optimisations simples sur un volume réduit de données, on peut graduellement enrichir nos outils en ajoutant de plus en plus de complexité au jeu d'apprentissage.

À partir des données et outils à notre disposition, il est possible de générer une infinité de réseaux différents : quels systèmes doivent être intégrés au jeu d'apprentissage ? Quelle sélection opérer ? Comment choisir leurs proportions relatives ? Combien de fonctions de symétrie utiliser ? Lesquelles ? Quelle forme donner au réseau de neurones ? Quels hyper-

paramètres choisir?

On trouve des pistes de réponses dans les sections précédentes, et l'idée de ces travaux n'est pas de présenter une étude exhaustive assurant d'obtenir **le** meilleur outil possible. En revanche, il semble pertinent d'en présenter plusieurs, et d'expliquer comment les uns justifient les autres : la probabilité d'obtenir un potentiel performant au premier essai est faible. Dès lors, les changements donnant lieu aux améliorations successives constituent un sujet d'étude intéressant, puisque chaque ajout, retrait ou ajustement est motivé par une intuition physique.

En particulier, discutons en amont de la présentation des apprentissages le choix du rayon de coupure des fonctions de symétrie utilisées pour décrire l'environnement atomique. Le choix de $r_{\rm cut}$ est le siège d'un compromis entre précision et transférabilité. Le potentiel répulsif exact - défini rigoureusement dans (3.12) - s'annule exactement au-delà des courtes distances. En augmentant trop $r_{\rm cut}$, on ne capture plus aucune information liée aux interactions répulsives, mais simplement les différences des parties électroniques entre *DFT* et *DFTB*. Notre objectif est de décrire l'énergie répulsive du système; il se trouve qu'on la calcule comme une différence d'énergie entre *DFT* et *DFTB* (6.2), mais cette différence en tant que telle **n'est pas** notre objectif final. Augmenter $r_{\rm cut}$ nous permettrait globalement d'être plus *précis* pour reproduire cette différence, mais n'aiderait pas à mieux décrire l'énergie répulsive et nuirait à la *transférabilité*.

Une première intuition pour un choix pratique fut de s'inspirer pour le Silicium du potentiel de types splines, obtenu par A. Sieck lors de sa thèse [36]. Le rayon de coupure utilisé par ces dernières $r_{\rm cut}^{\rm splines} = 2.54$ Å est assez pertinent lorsque l'on se concentre sur les contributions à deux corps en jeu pour les cristaux cubique simple et diamant du Silicium, employés pour optimiser les splines.

Néanmoins, ce choix s'avère trop limitant dans notre cas : d'une part la distance au plus proches voisins de certains systèmes étudiés lors de cette thèse (par exemple 2.50 Å pour le cristal β -Sn) avoisinent de très près cette valeur. D'autre part, l'une des forces des fonctions de symétrie que l'on emploie est la faculté des fonctions G_4 de tenir compte des variations angulaires au sein du système. La forme même des fonctions de symétrie angulaires (5.13) dépend explicitement des distances (r_{ij} et r_{ik}) de l'atome central *i* à ses voisins *j* et *k*, mais aussi de la distance entre les voisins *j* et *k* (r_{jk}). La distance r_{jk} peut excéder r_{cut} si celui-ci est trop petit, et l'application de la fonction de coupure (5.11) annule alors la contribution du triplet. Pour $r_{cut} < 3$ Å, on observe fortement ce phénomène.

A l'inverse, on note après quelques essais-erreurs sur des apprentissages tests que la précision est assez faiblement améliorée lorsque l'on augmente $r_{\rm cut}$ entre 3.0 Å et 3.3 Å : on choisit donc de fixer $r_{\rm cut} = 3$ Å pour la suite de ces travaux.

D'autre part, on utilise pour les applications comme celle-ci des réseaux de neurones de deux couches cachées, contenant quelques dizaines de neurones, et quelques dizaines de fonctions de symétrie. Si le nombre de neurones par couche cachées impacte évidemment la flexibilité, leur ajustement influe de manière modérée sur la précision finale : globalement, les trois facteurs déterminant l'ordre de grandeur de l'erreur finale sont la qualité du jeu de données de référence, les fonctions de symétrie utilisées et le choix de l'algorithme d'apprentissage.

Intéressons nous à la qualité du jeu de données de référence en présentant plusieurs réseaux pour lesquels le processus de sélection de fonctions de symétrie et l'algorithme d'apprentissage est identique (filtre de Kalman étendu - voir 6.2.2). Les fonctions de symétrie initiales sont les mêmes, on les pré-sélectionne suivant la plage de valeur qu'elles adoptent sur le jeu de données de référence puis elles sont élaguées par importance relative suite à plusieurs apprentissages successifs (voir annexe A). La principale différence entre ces réseaux est donc la constitution de leurs jeux de données de référence respectifs. On présente dans la suite l'entraînement de trois réseaux (nommés respectivement $\mathbf{R}, \mathbf{R}^{+\mathbf{L}}$ et $\mathbf{R}^{+\mathbf{L}+\mathbf{V}}$), paraissant pertinents pour expliquer quelques points d'intérêt. Le jeu de données de référence de chaque réseau est une version enrichie - et plus complexe - du précédent ⁴ et les proportions exactes de chaque sous-système peuvent varier.

7.2.1 Apprentissage sur des systèmes simples

Dans le cas du premier réseau R, ces données de référence sont les suivantes :

- dynamiques moléculaires pour les monocristaux (β -Sn, cubique Weaire-Phelan, diamant, hexagonal, monoclinique).
- dynamiques moléculaires pour les configurations amorphes (93, 96 et 100 atomes).
- calculs statiques pour les géométries de nano-clusters de [90], réoptimisées au préalable par *VASP*.
- dimère Si_2 pour lequel on fait varier la distance $r \in [0.8, 3.0]$.

Après optimisation des hyper-paramètres (voir 7.3), le réseau utilise 37 fonctions de symétrie (21 fonctions G_2 et 16 fonctions G_4), avec deux couches cachées de 24 neurones chacune.

Les fonctions de symétrie sont notamment sélectionnées automatiquement par importance relative (voir annexe A). On note qu'on emploie légèrement plus de fonctions de symétrie radiales parmi celles retenues; ceci s'explique par la forte proportion de mono-cristaux dans les données de référence, pour lesquels les angles varient assez peu. Autrement dit, les distances semblent ici relativement plus importantes que les angles, et le processus de sélection des fonctions de symétrie semble favoriser une forte sensibilité radiale. On présente l'évolution de l'erreur sur l'énergie répulsive et les forces répulsives au cours de l'entraînement dans la figure 7.4.

On observe une assez bonne capacité d'optimisation pour les erreurs respectives sur les énergies et les forces répulsives; au meilleur du compromis entre ces deux grandeurs, on atteint les précisions présentées dans la table 7.1.

Remarque : la racine de l'erreur quadratique moyenne (*root mean square error* - **RMSE**) ou l'erreur moyenne absolue (*mean absolute error* - **MAE**) sont les principales métriques

^{4.} Pour des raisons logistiques de calendrier, certaines données de référence n'étaient pas disponibles lors de l'entraînement des premiers réseaux de neurones présentés.



FIGURE 7.4 – Processus d'apprentissage : réseau global \mathbf{R} avec un jeu de données de référence d'une complexité modérée.

Grandeur	<i>RMSE.</i> jeu d'apprentissage	RMSE.jeu de test
Énergies	22.012 meV/atome	23.232 meV/atome
Forces	0.340 eV/Å	$0.351 \ \mathrm{eV/\AA}$

TABLE 7.1 – Erreurs sur les énergies et les forces : réseau global \mathbf{R} avec un jeu de données de référence d'une complexité modérée.

présentées pour juger de la qualité de ce genre de méthodes. Ceci se justifie plutôt bien : il est très aisé de les calculer à la volée durant l'entraînement d'un réseau de neurones, et elles permettent de juger avec un simple nombre leur performance globale.

Ces erreurs sont relativement satisfaisantes, bien que plus élevées que les précisions accessibles pour un potentiel numérique global au Silicium [98] ($\approx 10 \text{ meV}/\text{atome et} \approx 0.2 \text{ eV}/\text{Å}$). Le rayon de coupure de notre application étant bien plus réduit (3 Å contre 5.5 Å), une perte de précision est attendue : on ne veut reproduire que les contributions répulsives, et non corriger les approximations longue portée de *DFTB*. Ces données de référence sont relativement basiques, puisqu'il ne s'agit que de systèmes simples (Si_2), stables (nano-clusters) ou légèrement hors équilibre (dynamiques moléculaires); il est cependant encourageant de noter qu'on est effectivement capable de les reproduire efficacement.

7.2.2 Ajout de configurations liquides

Afin d'appréhender la capacité de nos outils à étudier une plus grande variabilité spatiale, intéressons nous à l'optimisation du deuxième réseau de neurones $\mathbf{R}^{+\mathbf{L}}$. Le jeu de données de celui-ci contient les mêmes systèmes que \mathbf{R} (dynamiques moléculaires pour les monocristaux et les amorphes, dimère Si_2 et géométries optimisées nano-clusters) dans des proportions légèrement différentes, auxquelles on ajoute des configurations liquides issues de dynamiques moléculaires à très haute température ($T_{\mathbf{lig}} = 3000 \text{ K}$).

Après optimisation des hyper-paramètres, le réseau emploie 35 fonctions de symétrie (8 fonctions G_2 et 27 fonctions G_4), avec deux couches cachées de 30 neurones chacune. On note cette fois-ci que la complexité accrue apportée par les configurations liquides s'accompagne de l'augmentation assez importante du nombre de fonctions de symétrie angulaires présentes par rapport au réseau **R**. Un aperçu de l'entraînement de ce réseau est disponible dans la figure 7.5.



FIGURE 7.5 – Processus d'apprentissage : réseau global $\mathbf{R}^{+\mathbf{L}}$ avec un jeu de données de référence d'une complexité intermédiaire.

Là encore, l'optimisation converge assez rapidement vers des plateaux pour les énergies et forces répulsives. Le meilleur compromis nous semble avoir lieu vers l'époque 43, et les erreurs respectives sont présentées ci-dessous (Table 7.2).

L'inclusion de liquides provoque une augmentation sensible de l'erreur commise sur les énergies répulsives. Cette augmentation modérée n'est pas une surprise en soi : la variabilité configurationnelle des liquides est très importante. Par ailleurs, ces systèmes sont assez éloignés des configurations pour lesquelles les approximations de DFTB sont

Grandeur	<i>RMSE.</i> jeu d'apprentissage	RMSE.jeu de test
Énergies	36.146 meV/atome	41.097 meV/atome
Forces	$0.383 \ \mathrm{eV/\AA}$	0.349 eV/Å

TABLE 7.2 – Erreurs sur les énergies et les forces : réseau global $\mathbf{R}^{+\mathbf{L}}$ avec un jeu de données de référence d'une complexité intermédiaire.

optimales : il est fréquent que les environnements proches de chaque atome soient assez singuliers, avec un nombre de liaisons par atome très variable. Il est donc tout à fait probable que les contributions électroniques prédites par VASP et DFTB soient assez différentes dans ce cas. Cette assertion est partiellement attestée par la difficulté que l'on observe en pratique à converger les calculs auto-cohérents de E_{coul} (voir 3.2) dans le cas des liquides. La charge de travail incombant au réseau est dès lors alourdie; l'augmentation de l'erreur observée semble en être une conséquence visible. Plus étonnamment, la qualité de notre représentation des forces interatomiques n'est pas amoindrie. Cette bonne surprise est complexe à pleinement interpréter, mais parait indiquer que nos outils demeurent très capables de supporter la complexité supplémentaire liée aux configurations liquides.

Jusque là, l'indicateur (RMSE) utilisé pour juger de la précision des potentiels répulsifs numériques présentés semble confirmer la fiabilité de nos outils. Néanmoins, une seule valeur quantitative peut ne pas rendre compte dans le détail de la précision d'un réseau de neurones couvrant une large variété de systèmes. C'est en tout cas le constat sans appel effectué au cours des travaux de cette thèse. Par exemple, la prédiction des courbes énergie vs. volume des cristaux est relativement médiocre (Fig. 7.6).

On note une forme globalement mauvaise : les minima d'énergie semblent prédits à des volumes presque comparables à VASP, mais on dénombre de nombreuses fluctuations inquiétantes, surtout aux faibles volumes - et donc au courtes distances. L'exemple le plus frappant - et le plus rédhibitoire - est observé pour le cristal β -Sn autour de 11 Å³/atome avec un puits de potentiel très loin de l'équilibre. On comprend bien pourquoi ce genre de comportement est assez préjudiciable - en encourageant le rapprochement excessif entre les atomes - d'autant que la représentation des faibles distances interatomiques est capitale pour une interaction aussi localisée que celle due aux potentiels répulsifs. On peut en déduire que les dynamiques moléculaires seules ne permettent pas de couvrir l'intégralité des distances d'intérêt.

7.2.3 Ajout des courbes énergies vs. volume

Cette partie présente l'entraînement du réseau de neurones utilisé dans le chapitre 8.

Dans l'idée de remédier aux difficultés notées dans la partie précédente, on entraîne un nouveau réseau de neurones, $\mathbf{R}^{\mathbf{L}+\mathbf{V}}$, avec un jeu de données de référence aspirant à pallier ce problème. En plus des mêmes sous-systèmes que les réseaux précédents (dynamiques moléculaires pour les monocristaux, les liquides et les amorphes, mais aussi géométries optimisées des clusters et dimères Si_2) on inclut également des courbes énergies vs. volume pour les monocristaux (comme dans la figure 7.6) et des géométries légère-

Stabilité relative des cristaux de Silicium pur



FIGURE 7.6 – Prédiction des énergies de liaison en fonction du volume pour les cristaux de Silicium pur par un réseau $\mathbf{R}^{+\mathbf{L}}$ inadéquat.

ment compressées ou étirées pour les nano-clusters ⁵ (voir 7.1.2.1). Après optimisation des hyper-paramètres, on obtient un réseau avec 31 fonctions de symétrie (23 fonctions G_2 et 8 fonctions G_4), et deux couches cachées de 20 neurones chacune. Ici, on remarque à nouveau que la présélection favorise proportionnellement les fonctions de symétrie radiales. L'ajout de systèmes compressés/étirés s'accompagne donc d'une importance décrue de la variabilité angulaire : ces nouvelles données de référence ne présentent littéralement **aucune** variabilité angulaire, ce phénomène n'est donc, somme toute, pas surprenant.

Comme toute inclusion de nouveaux systèmes au sein du jeu de données, l'ajout de courbes *énergie vs. volume* vient avec son lot de problèmes nécessitant une attention particulière. Ceux-ci s'expliquent tous en constatant d'une part que pour de telles données les énergies et forces répulsives de référence peuvent être très fortes, supérieures de plusieurs ordres magnitudes à celles rencontrées jusqu'alors. D'autre part, leur greffe n'aspire qu'à assurer une allure globale correcte (entre autre, pas d'attraction à courte portée), sans

^{5.} Pour les nano-clusters, l'idée est moins de représenter exactement les géométries hors équilibre que d'échantillonner le paysage énergétique autour des configurations minimales pour améliorer la prédiction de forces dans cette région de l'espace des configurations.

que l'on nécessite dans les faits une précision extrême sur ces données.

Nos outils s'optimisent en minimisant une fonction coût (voir 6.2) sommant l'erreur sur l'intégralité du jeu de données d'apprentissage. En conséquence, une erreur sensible sur des données aux énergies et forces élevées est bien plus pénalisante qu'un erreur relative similaire pour des données à l'équilibre, où les énergies et les forces sont faibles. Dès lors, si l'on ajoute trop d'énergies et forces répulsives de référence imposantes, le réseau s'optimise naturellement en étant très précis sur celles-ci au détriment des données à l'équilibre pour lesquelles une quasi-exactitude est capitale. Afin de contourner ces écueils, il est de mise de sélectionner avec précaution les boites compressées que l'on inclut au jeu d'apprentissage : conceptuellement, il n'en faut ni trop pour ne pas surcharger les données, ni trop peu pour assurer une allure physiquement acceptable pour l'intégralité des données.

Pour les nano-clusters, ceci est relativement aisé : on ne veut qu'explorer le paysage d'énergie potentielle proche des équilibres locaux. On peut donc se contenter de quelques configurations peu modifiées lorsque l'on intègre les nano-clusters compressés ou étirés, absents jusqu'alors, au données d'apprentissage. Pour les cristaux, on choisit une approche plus systématique : on conserve l'intégralité des points autour du minimum d'énergie, et l'on en sélectionne de moins en moins à mesure que l'on s'éloigne de l'équilibre. Par ailleurs, on décide pour la même raison de réduire légèrement le nombre de dimères inclus au jeu de données.

On présente dans la figure 7.7 l'évolution des erreurs sur les énergies et forces répulsives au cours du processus d'apprentissage.

L'ajout de configurations hors-équilibre, avec des forces et des énergies importantes, fait comme prévu augmenter significativement les racines de l'erreur quadratique moyenne relevées au long de l'apprentissage. On cherche à atteindre le meilleur compromis possible et on arrive au final à la précision présentées dans la table 7.3.

Grandeur	<i>RMSE.</i> jeu d'apprentissage	RMSE.jeu de test
Énergies	0.961 eV/atome	0.599 eV/atome
Forces	1.057 eV/Å	$1.279~{\rm eV/\AA}$

TABLE 7.3 – Erreurs sur les énergies et les forces : réseau global $\mathbf{R}^{+\mathbf{L}+\mathbf{V}}$ avec un jeu de données de référence d'une complexité élevée.

Ces erreurs sont élevées par rapport à celles (\mathbf{R} et $\mathbf{R}^{+\mathbf{L}}$) présentées précédemment, et méritent une étude approfondie. L'augmentation forte de la *RMSE* sur les énergies est la manifestation de deux phénomènes : d'une part, le réseau peine majoritairement à reproduire avec précision l'énergie des systèmes lorsque celle-ci élevée. On observe un semblant de palier variant pour chaque système au dessus duquel on sous-estime fortement les énergies par rapport à la référence. Ces systèmes étant très loin de l'équilibre, ce défaut est acceptable.

Suite à l'introduction des courbes *énergies vs. volume* au sein des données d'apprentissages, on observe une dégradation de la qualité de prédiction sur les autres systèmes, et notamment les dynamiques moléculaires. En un sens, ces dernières sont masquées par la



FIGURE 7.7 – Processus d'apprentissage : réseau global $\mathbf{R}^{+\mathbf{L}+\mathbf{V}}$ avec un jeu de données de référence d'une complexité élevée.

présence des nouvelles donnes. Les courbes énergie vs. volume suivent chacune une direction spécifiques de l'espace des configurations, et le réseau semble logiquement privilégier une précision accrue dans ces directions en raison des énergies et forces élevées qu'elles présentent. Les dynamiques moléculaires explorent en principe l'intégralité des directions accessibles aux systèmes, mais de manière moins extrême. Le réseau leur accorde une importance moindre au profit des courbes énergies vs. volume. Néanmoins, la hiérarchie énergétique des différentes phases reste respectée (voir la figure 8.1 dans le chapitre suivant) et la prédiction des courbes énergie vs. volume est très améliorée (voir la figure 8.2 dans le chapitre suivant)

Comme nous l'avons vu plus haut (fig. 7.6), les métriques RMSE et MAE s'avèrent intéressantes pour une première évaluation rapide mais sont incomplètes. Ces mesures d'erreurs sont très sensibles à des modifications minimes du jeu de données d'apprentissage, mais leur minimisation ne constitue pas une fin en soi. En retirant certaines configurations des données d'apprentissage - notamment et surtout des points des courbes énergie vs. volume - on obtient aisément des RMSE. de l'ordre de $1 \cdot 10^{-1}$ eV/atome pour les énergies répulsives et $2 \cdot 10^{-1}$ eV/Å pour les forces. De tels réseaux voient néanmoins leur qualité de représentation être moindre : nombre de grandeurs étudiées plus loin telles que les densités d'états vibrationnels - sont au final très mal décrites, indiquant des forces interatomiques assez mal évaluées. Au bout du compte, notre objectif est, plus que de parvenir à des erreurs impressionamment faibles, de reproduire convenablement des comportements clés du Silicium. On conserve donc pour la suite le réseau présenté dans la figure 7.7, en dépit de l'erreur quadratique moyenne observée, et on étudie dans le chapitre 8 nombre de propriétés nous semblant importantes pour discuter de sa performance *in-situ*.

7.3 Résumé et procédure à suivre

Avant d'avancer vers l'étude des performances du réseau de neurones obtenu précédemment, il semble important de conclure les quelques pages précédentes en présentant une *marche à suivre* pertinente pour générer efficacement un potentiel répulsif numérique. Cette démarche *ad-hoc* résume les points rencontrés lors des présents travaux : elle doit être vue comme une somme d'expérience accumulée par processus d'essai-erreur et n'a pas prétention d'être une vérité scientifique. La qualité des justifications de chaque choix peut être variable et n'a pas toujours été testé extensivement ni soumise à la critique d'experts du domaine. Ceci dit, cette expérience n'est pas sans valeur pour quiconque souhaite travailler à des applications similaires.

Le point capital de tout travail reposant sur des méthodes d'apprentissage automatique est également leur point de départ : le jeu de données de référence. Celui-ci doit être représentatif de l'objectif recherché; il n'est en aucun cas nécessaire de viser une grande capacité de généralisation si l'application ne le requiert pas. C'est en grande partie de la volonté de couvrir une majorité des systèmes du Silicium pur qu'apparaissent les difficultés rencontrées dans ces travaux.

Dans l'idéal, le jeu de données utilisé doit contenir plusieurs instantanés décorrélés issus de dynamiques moléculaires *DFT*. Chacune de celles-ci doit-être à température élevée pour une meilleure exploration de l'espace des configurations, sans pour autant excéder la température de fusion du système (sauf évidemment si l'on veut étudier des liquides). Enfin, il est également souhaitable - mais pas indispensable - d'ajouter quelques courbes énergies vs. volume. Celles-ci doivent être extrêmement minoritaires car les fortes énergies et forces qu'elles présentent compliquent légèrement le processus d'apprentissage. Elles sont néanmoins pertinentes dans le cas du potentiel répulsif à très courte portée pour prévenir le rapprochement excessif d'atomes. Par ailleurs, on les obtient facilement avec VASP, par quelques calculs statiques en modifiant simplement le facteur d'échelle des boîtes de simulation (paramètre sur la seconde ligne du fichier d'entrée POSCAR).

Il nous faut ensuite effectuer, pour chacune des boîtes sélectionnées, un calcul statique DFTB sans énergie répulsive. Si un apprentissage sur les forces répulsives est désiré, il faut activer le mot-clé *CalculateForces* dans le fichier d'entrée de DFTB+. L'efficacité numérique de cette méthode semi-empirique permet d'effectuer l'intégralité des calculs en quelques heures tout au plus sur un ordinateur de bureau. Pour les systèmes périodiques, l'emploi de la même grille de points K utilisée avec VASP est recommandée. De plus, il est important d'identifier les éventuels systèmes pour lesquels le calcul statique autocohérent n'aurait pas convergé et de les exclure de la base de données d'apprentissage. Enfin, appliquer une translation telle que présentée précédemment (voir 7.1.3) n'est pas

nécessaire en soi mais facilite beaucoup la comparaison avec d'autres méthodes.

On convertit ensuite les données de référence dans un format supporté par l'outil d'optimisation que l'on souhaite utiliser (N2P2 dans le cas des présents travaux). Avant de procéder à un quelconque apprentissage, nous avons néanmoins besoin d'un jeu de fonctions de symétrie nous permettant de décrire les environnements atomiques. Cette étape demande, *a minima*, de connaître le rayon de coupure au-delà duquel les contributions s'annulent. Celui-ci peut être identique à celui employé pour d'éventuels potentiels répulsifs de paire existants, mais il est en pratique plus intéressant de le choisir de manière à couvrir les premiers voisins de tous les systèmes étudiés. Nous avons vu (voir 3.3) qu'il est préjudiciable d'étendre beaucoup plus la portée des interactions répulsives.

On génère ensuite (voir A.1) un jeu conséquent (typiquement plusieurs centaines) de fonctions de symétrie radiales ou angulaires. Pour les fonctions G_2 , il est parfois intéressant d'avoir également un sous-jeu obtenu pour un rayon de coupure plus petit. Pour les fonctions G_4 , cette démarche est en pratique assez vaine : la forme de ces fonctions dépend des trois distances r_{ij} , r_{ik} et r_{jk} de chaque triplet d'atome ijk (5.13). Avec un rayon de coupure trop petit, la troisième distance est bien souvent supérieure à r_{cut} et la contribution du triplet est nulle. Par ailleurs, au sein de N2P2, la présence d'atomes sans voisins dans le jeu d'apprentissage entraîne l'apparition de matrices nulles lors de l'emploi du filtre de Kalman étendu (voir 6.2.2) : ces matrices nulles entraînent l'échec systématique des inversions matricielles propres à l'algorithme.

Faisons le point : nous avons pour le moment à notre disposition un jeu de données convenablement sélectionnées contenant positions atomiques (et d'éventuels paramètres de maille), énergie répulsive totale de référence, forces répulsives interatomiques de référence, mais aussi un large jeu de paramètres de fonctions de symétrie radiales et angulaires. Dans un premier temps, on peut pré-sélectionner certaines de ces fonctions de symétrie en tenant compte de la plage de valeurs qu'elles adoptent sur les configurations de référence (voir A.2). Dans l'idée, cette pré-sélection permet de conserver un nombre N_{ACSF}^{init} plus réduit de fonctions, typiquement une centaine; on souhaite dès lors ne retenir parmi celles-ci que les plus importantes pour l'apprentissage.

Après avoir effectué les normalisations (préalables dans le cas de N2P2 - voir annexe C.1.2.2) recommandées, on entraîne⁶ dans un premier temps un réseau de neurones relativement massif (typiquement de forme N_{ACSF}^{init} - 50 - 50 - 1, c'est-à-dire avec N_{ACSF}^{init} neurones d'entrée, deux couches cachées de 50 neurones et un neurone pour la couche de sortie). Il n'est pas requis d'effectuer un nombre d'époques d'apprentissage trop conséquent, et quelques dizaines sont en pratique suffisantes. Ce premier entraînement est en général assez long à cause du nombre important de fonctions de symétrie, et on peut accélérer le processus en n'optimisant le réseau que sur les énergies répulsives sans que cela ne soit trop dommageable. À l'issue de ce premier apprentissage, on sélectionne l'époque avec l'erreur globale la plus faible; on est alors à même de supprimer (voir A.3) les fonctions de symétrie ayant un impact relativement mineur sur la valeur prédite au final.

^{6.} Il est recommandé d'employer le filtre de Kalman étendu, présent dans N2P2 et semblant plus performant que la descente de gradient. Les paramètres de l'algorithme sont ardus à ajuster : il est préférable d'utiliser leur valeur par défaut.

On réitère si besoin cette boucle [$apprentissage + \acute{e}lagage$] jusqu'à avoir entre 25 et 50 fonctions de symétrie.

On a alors notre jeu final de N_{ACSF} fonctions de symétrie; il ne reste plus qu'à obtenir notre réseau final. Pour cela, un dernier ajustement est requis : la détermination du nombre de neurones par couche cachée, que nous recommandons d'être identique pour les deux couches du réseau. En effet, nous n'avons pas observé dans ces travaux d'impact notable dans le cas où les deux couches cachées n'ont pas le même nombre de neurones. Par ailleurs, l'utilisation d'une unique couche cachée a dans notre cas montré certaines limitations en terme de flexibilité, et l'apport de trois couches cachées n'a pas été notable : comme la majorité des applications similaires, nous préconisons d'en utiliser deux. Pour choisir le nombre de neurones par couche cachée, on passe là encore par un processus d'essai-erreur. Nous recommandons d'entraîner (sur l'énergie et les forces répulsives) plusieurs copies d'un même réseau pour lesquelles on fait varier le nombre de neurones par couche cachées régulièrement (typiquement de 5 en 5) entre 5 et N_{ACSF} neurones. Le nombre d'époque d'apprentissage peut être assez réduit, de l'ordre de la dizaine. On identifie alors la (ou les) forme(s) présentant une erreur moindre. Pour celle(s)-ci, on peut effectuer un apprentissage plus long, de l'ordre d'au moins 50 époques, et au plus quelques centaines.

On identifie enfin l'époque avec le meilleur compromis entre précision sur l'énergie et les forces répulsives pour les jeux d'apprentissage et de test : il s'agit de notre réseau final⁷ prêt à être greffé dans un fichier de *Slater-Koster* (voir C.2) et utilisé dans DFTB+!

^{7.} On peut si besoin déterminer *a posteriori* la répulsion de paire *ad-hoc* aux très courtes distances (voir 7.1.5).

Chapitre 8

Apport des potentiels répulsifs numériques

Après avoir créé, entraîné et sélectionné (Chap. 7) des réseaux de neurones aptes à reproduire la partie répulsive de DFTB pour le Silicium, nous voulons quantifier les gains de précision associés par rapport aux potentiels répulsifs de paire et à d'autres méthodes. En principe, les potentiels de paires pourraient se révéler peu transférables à des systèmes trop variés. Le passage vers des interactions répulsives à plusieurs corps semble pouvoir résoudre ce problème, mais qu'en est-il en pratique?

Nous nous attachons à présenter dans ce chapitre les résultats obtenus pour une large variété de systèmes de Silicium pur. Pour chacun d'entre-eux, il est important d'identifier des grandeurs physiques significatives, celles-ci pouvant évidemment être très diverses en fonction de l'application visée. Néanmoins, les potentiels répulsifs sont avant tout vus au sein du formalisme comme liant énergies et positions atomiques, car ils ne dépendent que de la densité électronique de référence $n_0(\vec{r})$ et non des spécificités de chaque système induisant un transfert de charge $\delta n(r)$ (voir 3.3). Aussi, on s'attelle principalement à caractériser les grandeurs liant structures et énergies, et on ne s'intéresse pas aux propriétés électroniques.

Ce chapitre est construit en trois parties, chacune orientée vers une classe de systèmes. Tout d'abord, nous détaillons l'apport des potentiels répulsifs numériques pour plusieurs cristaux de Silicium pur, avant de porter notre attention sur des objets isolés : les nanoclusters de Silicium. Enfin, nous nous intéressons aux phases désordonnées, c'est-à-dire aux liquides et aux amorphes. Cette liste ne couvre évidemment pas l'intégralité des systèmes possibles, mais présente une variété suffisante pour discuter des apports et limitations de cette approche de manière approfondie.

Pour chaque propriété étudiée, on propose de comparer les résultats obtenus avec notre approche (nommée **NN-E**_{rep} dans les figures) aux potentiels répulsifs de paire (**splines**). Les splines utilisées ici sont issues de [36]; ce choix nous permet de conserver les mêmes paramètres pour les termes électroniques de l'énergie *DFTB*, facilitant énormément la comparaison entre les deux potentiels répulsifs. On confronte également nos résultats à la méthode premier principe de référence (**VASP**) en conservant exactement les paramètres
utilisés pour générer les données de référence de notre outil NN- $E_{\rm rep}$ (GGA / PAW / ENCUT = 400 eV / EDIFF = 0.1 meV). Enfin, nous étudions la précision relative à deux approches purement paramétriques interfacées au code de dynamique moléculaire classique LAMMPS. Ces deux outils aspirent à reproduire l'énergie totale *DFT* et les forces interatomiques, à l'inverse de nos travaux focalisés sur un terme de *DFTB*, l'énergie répulsive.

Le premier de ces outils, nommé **NNP** dans les figures, utilise le même formalisme que le nôtre (fonctions de symétrie centrées sur les atomes de Behler-Parrinello et réseaux de neurones de haute dimension). Il a été obtenu grâce aux données de [98] : les boites de référence et les fonctions de symétrie sont les mêmes que celles fournis pour le Silicium pur dans l'article, et nous n'avons fait qu'effectuer le processus d'entraînement du réseau. Comme cet outil aspire à reproduire directement les énergies et les forces DFT, le rayon de coupure $(r_{cut}^{NNP} = 5.5 \text{ Å})$ est supérieur à ceux employés pour les potentiels répulsifs. L'article compare et discute des performances relatives de plusieurs approches par apprentissage automatiques dans la prédiction des propriétés de plusieurs éléments, dont le Silicium. Le jeu de données d'apprentissage comprend : cristal le plus stable (diamant), monocristaux (cubique centré, cubique face centrée, et diamant) sous contraintes graduelles, différentes surfaces, dynamiques moléculaires à différentes températures pour chaque monocristal (avec quelques températures au-delà du point de fusion) incluant ou non un défaut lacunaire cristallin. Les calculs sont également menés avec VASP, et les paramètres sont très similaires aux nôtres, à l'exception de l'énergie de coupure de la base d'ondes planes à 520 eV (contre 400 eV dans notre cas). Les auteurs présentent principalement les résultats pour de nombreux monocristaux, en étudiant leurs propriétés structurelles et énergétiques. De toutes les méthodes d'apprentissage automatique étudiées dans cet article, ce potentiel présente souvent les erreurs les plus marquées; ceci s'explique par la nature de l'investigation menée, qui aspire à utiliser un jeu de référence le plus petit possible. Les potentiels de type réseau de neurones ont un coût de calcul indépendant de la taille du jeu de données d'apprentissage (voir 5.3.1), et le potentiel NNP est donc pénalisé par ce choix. Sa précision reste tout de même largement acceptable, et aurait été aisément améliorable en enrichissant ses données de référence.

Le second potentiel purement numérique utilisé, nommé **GAP** [74] dans les figures, emploie quant à lui une approche différente : les méthodes à noyaux (présentées plus tôt, voir 5.3.2). Le détail de l'obtention de ce potentiel numérique de portée générale pour le Silicium ainsi que les paramètres de celui-ci sont disponibles dans la publication originale [74]. Cet outil est construit dans l'objectif de représenter l'intégralité des phases du Silicium pur, et est entraîné sur une gamme de systèmes très large : atome isolé, monocristaux (diamant, β -Sn, hexagonal simple, hexagonal diamant, cubique centré *bcc* et *bc8*, cubique face centrée, hexagonal compact, tétrahédrique *st12*), liquides, amorphes, surfaces du diamant, (bi-)lacunes et interstitiels du diamant, dislocations ou encore différentes hybridations (*sp* et *sp*²). Tous ces résultats ne sont pas obtenus avec *VASP*, mais avec *CASTEP* [99], et les paramètres sont sensiblement différents de ceux utilisés pour générer notre potentiel : on note par exemple l'emploi de la fonctionnelle GGA *PW91* [19, 100] (contre *PBE* dans notre cas), le choix de l'énergie de coupure pour la base d'ondes planes utilisée à 250 eV (contre 400 eV dans notre cas) ou encore l'utilisation d'un pseudopotentiel ultra-doux (*ultrasoft pseudopotential*). Le rayon de coupure de cet outil est là aussi supérieur à celui employé dans nos travaux, $r_{\rm cut}^{\rm GAP} = 5$ Å. Le coût numérique est bien plus important que pour le potentiel *NNP*, mais cet outil est supposé bien plus précis, et se veut représentatif de beaucoup plus de configurations du Silicium pur. Lors du test présenté par les auteurs pour ce potentiel, on observe en pratique une très bonne reproduction des propriétés élastiques, vibrationnelles et thermiques, mais aussi de celles des défauts cristallins, de la dépendance énergétique par rapport au volume de la cellule unitaire, ou encore des distributions radiales et angulaires de plusieurs systèmes désordonnés. Ces résultats impressionnants justifient à eux-seuls l'emploi de ce potentiel, même à un prix numérique élevé et malgré l'absence d'indication, par exemple, quant à la précisions pour des systèmes non-périodiques.

8.1 Cristaux de Silicium pur

Dans cette section, nous présentons les apports des potentiels répulsifs à plusieurs corps utilisant des réseaux de neurones $(NN-E_{rep})$ pour plusieurs cristaux purs. Ceuxci sont les mêmes utilisés pour l'entraînement des potentiels : β -Sn, cubique (Weaire-Phelan), diamant, hexagonal et monoclinique. Ces symétries sont intéressantes à étudier car elles sont toutes relativement stables à différentes pressions, et sont parfois très proches les unes des autres dans le paysage énergétique.

8.1.1 Stabilité relative

Pour de tels systèmes, il est évidemment indispensable de prédire correctement leur stabilité relative. Il s'agit en effet d'une condition sine qua non avant d'investiguer plus en détail l'utilité de notre potentiel répulsif numérique : s'il n'est pas à même de discriminer convenablement parmi les systèmes employés pour l'entraîner, il semble assez vain de s'intéresser à sa capacité à reproduire des propriétés pour lesquelles il n'est pas directement construit. On compare dans la figure 8.1 les écarts d'énergies au Silicium diamant calculés par $NN-E_{rep}$ aux mêmes écarts prédits par : les splines, la méthode de référence VASP, les deux approches purement paramétriques (NNP et GAP). On confronte également nos résultats avec les valeurs disponibles sous la dénomination énergies de formation sur Materials Project.

Le procédé est identique pour chacun de ces calculs : on part de la géométrie de plus petite taille disponible sur *Materials Project* (2 atomes par cellule pour les symétries β -Sn et diamant, 46 atomes pour le cristal cubique, et respectivement 4 et 8 atomes pour les phase hexagonale et monoclinique). Pour chaque méthode (*VASP*, splines, *NN-E*_{rep}, *NNP* et *GAP*), on optimise la géométrie de chaque cristal sans modifier la forme de la cellule unitaire¹. On détermine ensuite l'énergie totale par atome de chaque cristal. Dans tous les cas, on observe que le Silicium diamant est le plus stable : on rapporte donc l'écart en énergie des quatre autres cristaux par rapport à lui.

^{1.} L'interfaçage des potentiels répulsifs de type réseau de neurones à DFTB+ ne nous permet pour l'instant pas de réaliser avec fiabilité ce type de calcul. On choisit donc, pour ne pas fausser la comparaison, de s'en affranchir pour toutes les méthodes étudiées.



FIGURE 8.1 – Écarts en énergie des différents cristaux au Silicium diamant.

On se réfère à *Materials Project* pour plusieurs raisons. D'une part, il s'agit de la source d'où proviennent les géométries initiales utilisées pour les monocristaux étudiés lors de cette thèse. De plus, le fonctionnement *communautaire* de cette plateforme permet d'avoir accès à des propriétés relativement fiables, avec notamment de nombreuses informations sur le détail des calculs effectués. Il s'agit d'ailleurs d'un point important des résultats présentés dans la figure : si la référence Materials Project est très similaire à nos calculs (fonctionnelle GGA, pseudo-potentiel PAW), on note de légères différences avec le jeu de données d'apprentissage du réseau de neurones $NN-E_{rep}$. D'une part, l'énergie de coupure des ondes planes utilisées dans le calcul VASP est légèrement différente : 520 eV contre 400 eV dans notre cas, pour les raisons évoquées plus haut (voir 7.1.2.1). D'autre part, les calculs de cette thèse (VASP, DFTB et différents potentiels purement numériques) sont effectués sans optimisation de la géométrie de la cellule unitaire. Ces deux variables expliquent la différence entre Materials Project et nos résultats VASP : l'amplitude négligeable des écarts observés confirme que notre choix de ENCUT = 400eV est acceptable, mais aussi que l'optimisation de la cellule unitaire n'a qu'un impact mineur sur la stabilité observée des monocristaux.

Pour le cristal β -Sn, les approches purement numériques (NNP et GAP) sont les plus performantes. Ceci s'explique aisément : la partie électronique du formalisme DFTB est assez peu adaptée pour représenter les configurations atomiques formant beaucoup de liaisons. La symétrie β -Sn voit ses atomes former six liaisons, et les potentiels répulsifs (splines et NN- E_{rep}) sous-estiment sensiblement la stabilité de cette phase.

Pour les autres monocristaux, la précision des potentiels répulsifs DFTB est bien plus satisfaisante, et est même très proche de la référence VASP pour les symétries hexagonale et monoclinique. Pour la phase cubique, on sous-estime assez fortement (d'un facteur deux) l'écart en énergie par rapport au diamant; cela reste néanmoins bien meilleur que pour les potentiels de type NNP et GAP. Ceux-ci sont aussi inaptes à représenter la phase monoclinique, complètement absente de leurs jeux d'apprentissage respectifs.

Au final, les potentiels répulsifs DFTB sont très similaires entre eux (le potentiel NN- $E_{\rm rep}$ semble à peine meilleur que les splines, mais aucun écart significatif n'est relevé) et reproduisent convenablement les résultats de la méthode de référence. Par ordre de stabilité décroissante, on a : (diamant,) hexagonal, cubique (Weaire-Phelan), monoclinique puis β -Sn. En revanche, les approches purement paramétriques sont moins fiables (GAPsemble légèrement moins mauvais que NNP); en particulier, l'erreur forte observée pour la phase monoclinique induit une mauvaise prédiction de sa stabilité relative, puisqu'elle est décrite par ces outils comme moins stable que la phase β -Sn.

8.1.2 Courbes énergie vs. volume

Les informations données par la seule stabilité relative des cristaux sont limitées aux minima d'énergie des différentes phases. Les potentiels répulsifs étant avant tout présentés comme une fonction de distances interatomiques, on se propose de tracer pour chacun de ces cristaux leur énergie de liaison en fonction du volume par atome. On rappelle que l'énergie de liaison est calculée comme la différence d'énergie entre le système à N atomes considéré et l'énergie de N atomes isolés :

$$E^{\text{liaison}} = E^{\text{tot}} - N \times E^{\text{iso}}.$$

Cette grandeur rend donc compte du gain en stabilité apporté par les liaisons formées par le système. Pour étudier sa dépendance au volume, on varie la taille des boîtes de simulation en mettant à l'échelle les paramètres de maille et les positions atomiques donc le volume et les distances. On peut comparer cette énergie calculée avec DFTB+pour les différents potentiels répulsifs ($NN-E_{rep}$ et splines) avec les énergies de liaisons obtenues avec VASP pour des systèmes identiques, comme dans la figure 8.2. On donne en annexe (E.1) ces mêmes courbes obtenues avec les approches purement paramétriques (NNP et GAP); les présenter ici nuirait fortement à la lisibilité de la figure.

Dans ce cas, l'apport du potentiel répulsif de type réseau de neurones $(NN-E_{rep})$ est bien plus notable. Les énergies de liaison prédites sont plus proches des références VASP, à l'exception peut-être du Silicium β -Sn pour lequel l'énergie de liaison aux courtes distances est plus surévaluée que pour le potentiel de paire : l'écart de qualité de représentation entre les deux potentiels répulsifs pour ce cristal est discutable. Ce résultat s'explique partiellement par l'incapacité méthodologique des approximations DFTB à reproduire les systèmes où les atomes forment un grand nombre de liaisons.

Quoi qu'il en soit, il est surprenant de noter que les volumes où les stabilités entre polymorphes s'inversent sont sensiblement mieux prédits par les splines. Par exemple, l'inversion de stabilité de la phase diamant vers la symétrie cubique est prédite à $V \approx 23.55 \text{ Å}^3/\text{atome par } VASP, V \approx 23.09 \text{ Å}^3/\text{atome par les splines et } V \approx 22.85 \text{ Å}^3/\text{atome par les splines}$



FIGURE 8.2 – Stabilité relative des différents cristaux de Silicium pur en fonction du volume pour DFTB avec différents potentiels répulsifs.

En revanche, la courbure des énergies de liaison est globalement améliorée en employant le potentiel répulsif à plusieurs corps. Ce phénomène est cohérent avec des observations dans la littérature [101] relatives à la difficulté de reproduire simultanément les variations d'énergie pour plusieurs systèmes trop différents avec un unique potentiel répulsif de paire. De plus, la meilleure allure globale des énergies semble confirmer que les réseaux de neurones permettent une meilleure représentation des forces interatomiques.

Pour les potentiels purement numériques (NNP et GAP - voir figure annexe E.1), qui sont également des potentiels à plusieurs corps, les courbes énergie vs. volume sont plus irrégulières. Dans le cas de GAP, on note tout d'abord une forme idéale aux faibles distances, avec des énergies de liaisons très élevées pour les cinq cristaux étudiés. Aux longues distances, on observe une tendance décroissante pour les polymorphes diamant et hexagonal; ce comportement est peu surprenant, car peu de données aussi loin de l'équilibre figurent dans le jeu d'apprentissage de GAP. Ainsi, et même si les énergies de liaisons sont systématiquement surestimées avec GAP par rapport à $VASP^2$, leur

^{2.} Il est très probable que cette surestimation soit liée à la méthode de référence utilisée, qui est dans le cas de GAP assez différente de celle présentée dans la figure.

évolution en fonction du volume par atome est assez satisfaisante.

Pour NNP, les valeurs exactes des énergies de liaison sont bien plus proches de VASP, en particulier autour des volumes d'équilibre. On relève néanmoins des comportements inquiétants lorsque l'on s'en éloigne; si la légère décroissance observée aux longues distances peut ne pas être trop pénalisante, les fluctuations observées aux très faibles distances, avec même des puits de potentiel pour tous les cristaux sauf β -Sn, est problématique. Ce genre de comportement, non-physique, autorise en théorie un rapprochement excessif d'atomes et peut donc s'avérer rédhibitoire.

L'étude de la dépendance de l'énergie de liaison au volume pour chacun des cristaux indique que les approches DFTB supplantent les outils purement numériques. Aux longues distances, les contributions électroniques à l'énergie DFTB (identiques dans les deux cas) sont dominantes et semblent assez robustes; les courtes distances sont dépendantes de l'interaction répulsive utilisée, et notre approche $NN-E_{\rm rep}$ paraît clairement plus performante que les splines. Les approches purement paramétriques (NNP et GAP) sont évidemment plus flexibles, mais le manque de données de référence aux volumes éloignés de l'équilibre les pénalise sensiblement. En particulier, NNP semble permettre un rapprochement excessif d'atomes.

Au passage, on peut extraire l'énergie répulsive des calculs effectués ci-dessus. On présente ainsi dans la figure 8.3 la dépendance de l'énergie répulsive à la distance aux premiers voisins pour les cinq monocristaux. On trace d'une part les énergies répulsives obtenues avec les splines et notre potentiel répulsif numérique $NN-E_{\rm rep}$, et d'autre part la référence que nos outils sont supposés reproduire. Cette dernière valeur est calculée comme la différence (6.2) et translatée d'après (7.4).



FIGURE 8.3 – Énergie répulsive des monocristaux en fonction de la distance au plus proche voisin.

Pour les splines, on observe sans surprise que la forme du potentiel est la même pour chaque polymorphe; il s'annule à $r_{\rm cut}^{\rm splines} = 2.54$ Å, et croît régulièrement à mesure que la distance diminue. Au global, cette forme est acceptable. Dans le cas du diamant en particulier, pour lequel les splines sont optimisées, la forme globale du potentiel est correcte jusqu'à $r_{\rm cut}^{\rm splines}$.

Pour le cristal β -Sn, mal reproduit par les splines, on note pour la référence une forme assez surprenante, globalement décroissante à mesure que la distance interatomique diminue. Ici, il apparaît clairement que la partie électronique de *DFTB* attribue une forte énergie au polymorphe β -Sn. Ce cristal est relativement dense, comme l'attestent les volumes de la figure 8.2, et démontre surtout une haute coordinence pour chaque atome - situation pour laquelle la partie électronique de *DFTB* est moins adaptée.

Le potentiel répulsif numérique $NN-E_{\rm rep}$ performe légèrement mieux que les splines pour la symétrie β -Sn, sans pour autant pouvoir être qualifié de satisfaisant. La diminution de l'énergie de référence aux très courtes distances n'est pas reproduite, au profit d'une augmentation en accord avec les références des autres cristaux. Pour ces derniers, le potentiel $NN-E_{\rm rep}$ est tout à fait raisonnable, même si on relève un certain manque de flexibilité empêchant de coller parfaitement aux références. On observe enfin pour les très courtes distances que la pente est sensiblement plus douce que pour la référence; cela traduit une mésestimation des forces répulsives dans ces régions très éloignées de l'équilibre. Aucun puits de potentiel à proprement parler n'est à déplorer, et on n'autorise à priori pas de rapprochement excessif d'atomes. Autrement dit, pour des systèmes très compressés la prédiction n'est pas exacte, mais cette lacune paraît plutôt inoffensive.

8.1.3 Densités d'états vibrationnels

On s'intéresse ensuite aux propriétés vibrationnelles des cristaux pour deux raisons : leur bonne représentation dépend directement de la qualité des forces interatomiques et joue un rôle important dans la prédiction de nombreuses grandeurs d'intérêt telles que par exemple la conductivité thermique ou la capacité calorifique. On présente (Fig. 8.4) les densités d'états vibrationnels prédites par les réseaux de neurones ($NN-E_{rep}$) ou les splines et les valeurs de référence calculées avec VASP. Ces résultats sont obtenus en utilisant le logiciel phonopy [102] : ce dernier permet de déterminer de nombreuses propriétés vibrationnelles pour les systèmes périodiques et est compatible avec VASP, DFTB+ et $LAMMPS^3$. On optimise tout d'abord la géométrie d'intérêt avec l'outil désiré (VASP, DFTB+ avec les différents potentiels répulsifs ou LAMMPS avec les différents potentiels numériques), puis on réplique la cellule dans l'espace avant de lui appliquer quelques déplacements choisis en fonction de la symétrie du système, et on calcule finalement la courbe de dispersion des phonons en plusieurs points de la première zone de Brillouin, que l'on intègre ensuite sur les fréquences pour obtenir la densité d'états vibrationnels.

Deux éléments clés déterminent la qualité des résultats obtenus. D'une part, la taille de la supercellule étudiée : plus on réplique la cellule dans l'espace avant d'appliquer les déplacements, plus on est à même de pouvoir considérer les contributions de longue por-

^{3.} Via l'interface python phonoLAMMPS.



FIGURE 8.4 – Densités d'états vibrationnels des cristaux de Silicium pur pour différents potentiels répulsifs DFTB.

tée aux propriétés vibrationnelles. Le nombre de répliques utilisées dépend du système étudié, et est déterminé par un processus d'essai-erreur interrompu lorsqu'augmenter la taille de la supercellule impacte peu les résultats observés⁴. D'autre part, recourir à un grand nombre de points de l'espace réciproque augmente significativement la résolution des courbes observées. De manière très analogue au point précédent, on augmente graduellement le nombre de points utilisé jusqu'à convergence : dans notre cas, nous observons qu'une grille *Monkhorst-Pack* [103] de $30 \times 30 \times 30$ points suffit pour tous les calculs présentés.

Ces figures font état d'un constat satisfaisant : les densités d'états vibrationnels des cristaux étudiés sont globalement bonnes, et le potentiel répulsif de type réseau de neurones $(NN-E_{rep})$ surpasse sensiblement le potentiel de paire (splines).

Pour le Silicium β -Sn (8.4a), les deux potentiels répulsifs sont relativement mauvais pour rendre compte des modes acoustiques ($\omega < 7.5$ THz) mais sont assez similaires, indiquant que les faibles fréquences semblent principalement régies par les termes électroniques de *DFTB*. La reproduction des fréquences plus élevées est sensiblement améliorée par l'emploi du potentiel répulsif numérique, sans pour autant être très fidèle aux données de référence. Enfin, le réseau de neurones NN- $E_{\rm rep}$ semble légèrement plus précis pour les fréquences intermédiaires.

Si l'on s'intéresse à la phase cubique (8.4b), notre outil semble ne prédire qu'une version adoucie des densités d'états obtenues avec les splines. Les amplitudes et les fréquences sont en effet très similaires, et il semble que le potentiel de paire soit à peine préférable à son homologue numérique, sans que l'écart ne soit réellement significatif. Notons en revanche les quelques fréquences négatives observées pour le réseau de neurones; celles-ci indiquent que le système observé n'est pas exactement à son minimum, sans que cela ne semble être effectivement impactant au regard des points présentés plus tôt.

La densité d'états vibrationnels du Silicium diamant (8.4c) est mieux prédite par le potentiel répulsif de type réseau de neurones. Ici encore, les modes acoustiques sont très similaires, et légèrement mieux reproduits par les splines. Pour toutes les fréquences intermédiaires et les modes optiques ($\omega > 7.5$ THz), le potentiel répulsif numérique NN- $E_{\rm rep}$ suit presque exactement la référence VASP. Il en va rigoureusement de même pour le Silicium hexagonal (8.4d) : là aussi, les basses fréquences sont très similaires pour les deux potentiels répulsifs et le réseau de neurones excelle pour les fréquences intermédiaires et élevées. Pour celles-ci, le gain par rapport aux splines est encore plus marqué que dans le cas du Silicium diamant.

Portons maintenant notre attention sur la densité d'états vibrationnels du Silicium monoclinique (8.4e). Ici, le constat précédent semble se confirmer définitivement : les modes acoustiques ne sont presque pas influencés par le potentiel répulsif utilisé. En revanche, les plus modes optiques semblent mieux représentés par le réseau de neurones $NN-E_{\rm rep}$. Si la quantité de pics nuit à la lisibilité globale, il semble néanmoins plus aisé de déterminer pour chacun d'entre eux une correspondance VASP-potentiel répulsif numérique; dans le même temps, les splines semblent mésestimer plus fortement l'amplitude

^{4.} Au final, les cristaux étudiés contiennent respectivement 250 (β -Sn), 368 (cubique), 128 (diamant), 256 (hexagonal) et 216 (monoclinique) atomes.

et la fréquence des modes observés par rapport à la référence.

On présente également en annexe (voir E.2) les densités d'états vibrationnels prédites pour ces mêmes monocristaux avec les deux potentiels purement numériques (NNP et GAP). Concernant les quelques systèmes pour lesquels les résultats observés ont un sens physique, c'est-à-dire les symétries diamant et hexagonale (figures annexes E.2c et E.2d), on remarque un comportement différent pour les deux approches.

D'une part, NNP prédit légèrement mieux les modes acoustiques que les deux potentiels répulsifs DFTB (splines et $NN-E_{rep}$), sans pour autant coller globalement à la référence VASP, en particulier pour les modes optiques.

D'autre part, le potentiel GAP est sans conteste excellent pour le Silicium diamant, reproduisant parfaitement la densité d'états vibrationnels de référence pour toutes les gammes de fréquence. Le résultat est légèrement plus mitigé pour la symétrie hexagonale, et les modes de basse fréquence sont assez mal prédits; pour ce système, les approches DFTB, et en particulier nos outils ($NN-E_{rep}$), semblent préférables.

Ainsi, l'étude des densité d'états vibrationnels des monocristaux nous permet de noter quelques caractéristiques intéressantes. La première est l'inefficacité de nos outils à améliorer la précision des modes acoustiques : ceux-ci semblent assez peu sensibles au potentiel répulsif employé, et paraissent donc dominés par les contributions électroniques à l'énergie DFTB. On remarque ensuite la meilleure précision, légère mais présente, de notre approche pour un potentiel répulsif plusieurs corps $(NN-E_{rep})$ par rapport aux potentiel répulsif de paire de type splines à partir de $\omega \sim 7.5$ THz et jusqu'aux modes optiques. Ce constat est maintenu - et même accentué - dans le cas du diamant, alors même que les splines ont également été optimisées sur cette symétrie. Il nous paraît raisonnable de voir ici une nouvelle confirmation de l'intérêt des outils d'apprentissage automatique, et des précisions accessibles à travers eux. Une dernière caractéristique mitige néanmoins ce constat, puisqu'on voit que pour les méthodes purement numériques (NNP et GAP), une flexibilité non-bridée est à double-tranchant. On est en mesure de mieux représenter les modes acoustiques, puisque ces approches purement paramétriques sont affranchies de termes électroniques comme ceux de DFTB. Cette trop grande souplesse autorise néanmoins des comportements aberrants, comme observé avec NNP et GAP pour les cristaux β -Sn, cubique et monoclinique (figures annexes E.2a, E.2b et E.2e).

8.1.4 Défauts ponctuels du Silicium diamant

Enfin, intéressons nous aux énergies de formation de quelques défauts ponctuels apparaissant dans le Silicium diamant. Les cristaux sont des systèmes périodiques, se répétant à l'infini dans les trois directions de l'espace; un défaut est simplement l'interruption localisée de cette symétrie. Cette rupture peut se faire le long d'une ligne de dislocation (on parle de défaut linéaire), sur un plan (comme dans le cas de joints de grain ou de défauts d'empilement - on parle de défaut planaire), ou même apparaître à l'échelle de volumes (par exemple dans le cas de surfaces ou de matériaux poreux). Le dernier cas, celui étudié ici, est le cas de défauts ponctuels où la symétrie infinie n'est brisée qu'en un unique point du cristal. Nous nous concentrons sur deux cas de figure courants : le cas de lacunes, où un atome manque et laisse un site vide, et celui d'interstitiels, où un atome s'insère dans un espace entre les nœuds du réseau.

La base de données d'apprentissage de notre outil $(NN-E_{\rm rep})$ ne contient aucune configuration de la sorte, et mesurer l'écart d'énergie induit par un défaut constitue une évaluation pertinente de la transférabilité du réseau de neurones à des configurations supposément inconnues. La première classe de défaut étudiée est une simple lacune, c'est-à-dire que l'on retire un atome de Silicium au sein de la maille diamant. Les autres sont des interstitiels, respectivement tétrahédrique (figure 8.5a) et hexagonal (figure 8.5b), où un atome de Silicium supplémentaire se loge entre les atomes déjà présents.





(a) Défaut interstitiel tétrahédrique dans la maille de Silicium diamant.

(b) Défaut interstitiel hexagonal dans la maille de Silicium diamant.

FIGURE 8.5 – Défauts interstitiels tétrahédrique et hexagonal du Silicium diamant.

Pour chacun des cas étudiés, on part d'une configuration large ($N^{\text{atomes}} = 216$) afin de pouvoir considérer le défaut comme ponctuel. Cette taille nous paraît raisonnable car elle est en accord celles employées dans la littérature [74, 104]. On ajoute ou retire ensuite un atome de manière à générer localement la configuration souhaitée. On souhaite connaître l'écart de stabilité entraîné par cette modification : on relaxe donc la géométrie du système, sans modifier les paramètres de maille. Ce choix trouve deux motivations; la première, assez pragmatique, réside dans l'incapacité présente de notre méthodologie $(DFTB+ \text{ avec } NN-E_{rep})$ à relaxer ces paramètres; la seconde trouve une explication plus fondamentale : on veut étudier l'impact de défauts très localisés sur un cristal infini. Toutefois, la boîte de simulation est nécessairement finie, et les différentes méthodes pallient cet écueil en la créant des images périodiques artificielles à l'infini. Dès lors, relaxer les vecteurs de la boîte équivaut à renforcer la modification causée à l'infini par le défaut. Il semble donc important de figer ces paramètres de maille, pour minimiser l'impact aux très longues portées d'un défaut ponctuel. Cela nous permet en somme de nous concentrer sur une relaxation localisée autour de la perturbation. On calcule finalement l'énergie de formation du défaut en prenant comme référence l'énergie d'un

atome au sein du cristal parfait [104–106], telle que :

$$E_{\text{formation}}^{\text{defaut}} = E_{\text{totale}}^{\text{defaut}} - N^{\text{atomes}} \times E_{\text{atome}}^{\text{diamant parfait}}$$

Ici, $E_{\text{totale}}^{\text{défaut}}$ est l'énergie totale de la boîte de simulation contenant le défaut après relaxation, $E_{\text{atome}}^{\text{diamant parfait}}$ est l'énergie par atome du Silicium diamant non perturbé. Pour obtenir ces valeurs, différentes pour chaque méthode étudiée, on part d'une cellule unitaire de diamant parfait optimisée par VASP (paramètre de maille de 5.47Å), et on relaxe les positions des atomes avec les différents potentiels. Enfin, N^{atomes} vaut respectivement 216 - 1 = 215 ou 216 + 1 = 217 atomes selon le type de défaut considéré. Les résultats du calcul de cette énergie de formation avec VASP, les potentiels répulsifs de paire (splines), de type réseau de neurones ($NN-E_{\text{rep}}$), et pour les deux potentiels purement paramétriques (NNP et GAP) sont donnés pour chacun de ces défauts dans la figure 8.6.



FIGURE 8.6 – Comparaison des énergies de formation de défauts ponctuels.

Indépendamment du potentiel répulsif, on note que DFTB (splines et $NN-E_{rep}$) tend à surévaluer globalement l'énergie de formation des défauts. Il est donc plausible que la grandeur étudiée soit dominée par les contributions électroniques à l'énergie totale DFTB. Néanmoins, on note que l'emploi de tel ou tel potentiel répulsif modifie sensiblement les valeurs obtenues. Pour les lacunes, cette surévaluation est légèrement moins marquée dans le cas des splines. Ce défaut est de loin le plus stable des trois avec DFTB pour les deux potentiels répulsifs, ce qui montre un désaccord avec VASP; on devine une tendance à préférer les systèmes peu coordinés inhérente au formalisme semi-empirique que le réseau de neurones ne semble aucunement corriger.

En portant notre attention sur les interstitiels, la surévaluation globale des énergies de formation par *DFTB* est encore plus marquée, quelle que soit l'interaction répulsive. Une différence de taille est néanmoins à souligner, puisque les deux potentiels répulsifs montrent un désaccord sur l'interstitiel le plus stable. Le potentiel de type splines favorise de très peu l'interstitiel tétrahédrique, avec un écart d'énergie de formation de 0.041 eV par rapport à l'interstitiel hexagonal. Ce dernier est en revanche largement plus stable que son homologue d'après $NN-E_{rep}$ (écart d'énergie de 0.256 eV); ce résultat s'accorde avec VASP, où l'écart d'énergie est très similaire (0.233 eV).

La valeur absolue des énergies de formations de défauts ponctuels est donc très surestimée avec DFTB, qui pénalise d'autant plus les systèmes très coordinés. Pour ceux-ci, le potentiel numérique NN- $E_{\rm rep}$ s'avère au final préférable aux interactions de paires, sans pour autant corriger intégralement les diverses mésestimations de la méthode. L'amélioration de précision pour les interstitiels peut se justifier partiellement par la prise en compte de systèmes hautement coordinés dans le jeu de données d'apprentissage du réseau de neurones : le cristal β -Sn forme 6 liaisons, et de telles configurations apparaissent aussi localement au sein des systèmes désordonnés. En principe, ces derniers présentent également des atomes formant moins de 4 liaisons, mais ce phénomène est de toute évidence trop peu présent pour être assimilé : il serait donc intéressant d'étoffer nos données de référence avec diverses lacunes cristallines.

Par ailleurs, les potentiels purement numériques (NNP et GAP) prédisent des énergies de formation globalement comparables à VASP. En l'état, de telles configurations sont connues des potentiels, puisqu'une quantité variable de défauts cristallins figurent au sein de leurs jeux d'apprentissage respectifs. Dans le cas de NNP, le jeu d'apprentissage contient uniquement des lacunes; il est dès lors curieux de remarquer qu'il s'agit du défaut que ce potentiel reproduit le moins fidèlement. En particulier, la stabilité de l'interstitiel tétrahédrique est pour NNP presque identique à la référence VASP; l'erreur est sensiblement plus notable pour l'interstitiel hexagonal. Au final, la stabilité globale de ces trois défauts avec NNP est en désaccord avec la méthode premier principe : l'ordre prédit est le même que pour notre approche $NN-E_{rep}$.

Dans le cas de GAP, le jeu de référence contient une liste très importante de défauts cristallins, parmi lesquels figurent les lacunes et interstitiels étudiés. Cela se traduit effectivement dans cette figure, puisque les énergies de formation prédites s'approchent beaucoup de la référence VASP, et que la stabilité relative est celle attendue.

Pour résumer, les splines semblent être les moins à même de décrire efficacement les défauts ponctuels étudiés. On observe une amélioration sensible apportée par nos outils ($NN-E_{rep}$), notamment marquée pour les interstitiels; ce phénomène est néanmoins insuffisant, et les deux approches purement numériques surclassent DFTB. Le potentiel NNP est le moins bon des deux - ce qui s'explique partiellement par la teneur plus modeste en défauts de son jeu d'apprentissage. De son côté, GAP excelle sur cette propriété particulière pour laquelle une forte emphase est mise dans son jeu données de référence.

Ces quelques résultats semblent indiquer une plus grande capacité prédictive pour le potentiel répulsif de type réseau de neurones $(NN-E_{rep})$ lorsque l'on s'intéresse aux cristaux. Les écarts en énergies sont relativement similaires aux splines et surclassent NNP et GAP (fig. 8.1); l'étude comparative des courbes énergie vs. volume (fig. 8.2) et des densités d'états vibrationnels (fig. 8.4) renforce cette affirmation; enfin les énergies de formation de défauts ponctuels sont mieux décrites que pour les splines mais moins bien que pour NNP et GAP (fig. 8.6). Au global, notre outil montre des énergies et des forces interatomiques prédites plus proches de la méthode de référence que les splines et les approches purement numériques.

Notons que la meilleure prédiction du potentiel répulsif de type réseau de neurones sur les splines pour la reproduction des propriétés du Silicium diamant est intéressante : le potentiel de paire n'est optimisé «que» sur ce cristal et sur une phase cubique simple, alors qu'une beaucoup plus grande variété de systèmes apparaissent dans le jeu de données d'apprentissage construit lors de cette thèse. Ceci semble confirmer que le réseau de neurones est plus à même de supporter une complexité accrue et que la perte de précision associée est plus faible que pour les splines. Le potentiel de paire est tout de même pertinent sur bon nombre de systèmes, sans être optimisé explicitement pour cela. Même s'il n'est de toute évidence pas aussi précis que son équivalent numérique sur les exemples observés jusqu'à maintenant, sa robustesse est notable.

8.2 Nano-clusters de Silicium pur

Nous nous intéressons ici à la qualité de représentation des nano-clusters de Silicium pur compris entre Si₁₄ et Si₇₇. Les géométries étudiées proviennent de [90]; ce sont les mêmes utilisées pour générer le jeu de données d'apprentissage présenté précédemment (voir 7.1).

8.2.1 Accord global avec VASP dans la prédiction des configurations stables

Une première métrique intéressante consiste à évaluer la capacité des potentiels répulsifs à prédire le bon minimum pour une taille de nano-cluster donnée parmi plusieurs candidats. En effet, la recherche de structures de référence pour les systèmes non-périodiques est un processus long et assez coûteux : on génère avec plus ou moins d'intuition des configurations, dont il faut évaluer et comparer la stabilité relative. Pour des objets isolés dans l'espace, le nombre d'agencements atomiques possibles est colossal, et il s'avère souvent trop coûteux de mener une recherche exhaustive basée sur des méthodes premier principe telles que DFT. Une approche courante est d'effectuer une pré-sélection grossière avec des outils moins coûteux numériquement, tels que DFTB par exemple. Dans ce contexte, il est aisé d'imaginer comment la précision de l'outil en question est déterminante : plus celui-ci est fiable, plus la sélection appliquée peut-être agressive. Ceci permet de tester un plus grand nombre de candidats, améliorant d'autant la probabilité d'identifier **la** configuration la plus stable.

Dès lors, l'idée est la suivante : pour chaque taille de clusters on détermine - de différentes manières - le candidat le plus stable avec VASP parmi plusieurs configurations (entre 18 et 50 selon le cas). On effectue un calcul analogue avec DFTB (en employant les splines ou nos outils $NN-E_{rep}$) et avec des potentiels purement numériques (NNP et GAP), puis on évalue la concordance avec VASP.

Les géométries de nano-clusters sont, dans le cas de [90], des configurations obtenues par un algorithme évolutionniste (USPEX) [107,108] interfacé à DFTB+ pour l'évaluation de leur énergie, sélectionnées parmi les plus stables pour chaque taille de cluster et dont la géométrie est optimisée par un autre code DFT, FHI-aims [109]. Ce code, très adapté aux systèmes isolés, tient compte explicitement de tous les électrons du système, et les calculs sont réalisés avec la fonctionnelle GGA-PBE; par ailleurs, les fonctions de base ne sont pas des ondes planes (comme dans le cas de VASP) mais des orbitales numériques centrées sur les atomes.

Dans le cas des travaux de cette thèse, nous cherchons à comparer la précision de notre approche vis-à-vis de VASP. En conséquence, le jeu d'apprentissage de notre outil $NN-E_{\rm rep}$ contient ces mêmes configurations dont la géométrie a été réoptimisée par VASP. Ce sont vis-à-vis de celles-ci que se font toutes les comparaisons présentées dans la suite de cette section.



FIGURE 8.7 – Comparaison des accords avec DFT dans la prédiction des minima d'énergie de configurations de nano-clusters optimisées par VASP.

Dans un premier temps, on effectue un calcul statique avec les quatre méthodes étudiées (splines, $NN-E_{rep}$, NNP et GAP) sur les géométries de clusters optimisées avec VASP afin de déterminer leur faculté à s'accorder à la méthode de référence. On reporte pour cela la proportion de concordances exactes dans les minima prédits (P_{\min}) , et celle des cas où le cluster de plus basse énergie DFT est parmi les dix plus stables d'après chaque méthode (P_{top10}) . On observe dans ce cas que le potentiel développé lors de nos travaux (NN- $E_{\rm rep}$) est sensiblement plus performant que les splines et s'accorde globalement mieux avec VASP.

Pour les approches purement numériques (NNP et GAP), ne comprenant aucun système non périodique dans leur jeu

de données d'apprentissage, on observe d'une part que la concordance exacte avec VASP est plutôt rare, alors que dans le même temps la présence du minimum VASP parmi les 10 configurations prédites comme les plus stables est notoire - même meilleure que pour les

splines. Il semble donc, d'après leur P_{top10} élevé, que les potentiels numériques (NN- E_{rep} , NNP et GAP identifient efficacement des groupes de configurations dont les propriétés structurelles sont importantes pour la stabilité : par exemple la coordinence moyenne ou le nombre d'atomes dont l'environnement est quasi-cristallin, en fonction de la taille du cluster. Néanmoins, pour NNP et GAP, la détermination du minimum effectif parmi ces configurations semble presque aléatoire $(P_{\min}/P_{top10} \gtrsim 10\%)$. Ceci laisse deviner que la discrimination entre ces candidats pour les minima est relativement mauvaise.

Jusqu'ici, les indicateurs tendent à montrer que le potentiel répulsif de type réseau de neurones prédit assez bien les énergies des clusters étudiés. Intéressons nous ensuite à la qualité de représentation des forces interatomiques dans ces systèmes isolés. Pour cela, la démarche est la même que précédemment, mais à la place de calculs statiques on performe une optimisation de géométrie avec chaque outil (splines, $NN-E_{rep}$, NNP et GAP) avant de comparer les concordances P_{\min} et P_{top10} avec VASP.

On présente les résultats d'une telle démarche dans la figure 8.8. La différence avec les résultats précédents est assez frappante : tout d'abord, les approches purement numériques (NNP et GAP) performent de manière médiocre. Par ailleurs, les splines sont meilleures que leur homologue numérique $(NN-E_{rep})$. En optimisant la géométrie des clusters avec le potentiel répulsif de paire, les configurations de plus basses énergies semblent coïncider fréquemment avec celles obtenues par la même procédé avec VASP.

A l'inverse, les résultats pour notre approche sont plus mitigés : au global, la configuration de plus faible énergie coïncide perceptiblement moins souvent avec celle de VASP ($P_{\min}^{NN-E_{rep}} < P_{\min}^{splines}$), et les



FIGURE 8.8 – Comparaison des accords avec DFT dans l'optimisation et la prédiction des minima d'énergie.

dix meilleures géométries prédites par taille incluent également moins souvent la configu-ration la plus stable ($P_{\text{top10}}^{\text{NN-}E_{\text{rep}}} < P_{\text{top10}}^{\text{splines}}$).

Le processus d'optimisation de géométrie - faisant intervenir les forces interatomiques - semble aider grandement le potentiel répulsif de paire à converger vers les données de référence, et pénaliser dans le même temps le réseau de neurones $NN-E_{rep}$. Par ailleurs, il est important de souligner la plus grosse lacune de notre outil sur les systèmes isolés : il n'est pas toujours possible de converger vers une géométrie. À l'inverse des potentiels de paire convexes, convergeant par définition toujours quel que soit le système, rien ne contraint le potentiel numérique à présenter pour l'intégralité des configurations un minimum d'énergie aisément accessible dans son paysage d'énergie potentielle. Cela se traduit en pratique par des forces assez élevées, et le processus d'optimisation de géométrie est globalement moins robuste, nécessite bien plus d'itérations et peut ne pas aboutir.

Pour s'en convaincre, jetons un œil à la figure 8.9; celle-ci rend compte du processus d'optimisation de la géométrie de deux nano-clusters constitués de 14 atomes. Nous montrons les résultats pour ceux-ci car notre potentiel répulsif NN- $E_{\rm rep}$ parvient à optimiser la géométrie du premier (appelé Si_{14} n°1) mais échoue pour le second (appelé Si_{14} n°2). En revanche, les splines optimisent ces deux géométries sans problème. Avant de présenter les résultats en eux-même, définissons précisément les termes employés. L'objectif d'une optimisation de géométrie est d'atteindre des positions pour lesquelles les forces interatomiques sont les plus faibles possibles, et l'énergie totale ($E_{\rm tot}$ dans la figure) est alors en principe dans un minimum local.

Avec DFTB+, le processus s'effectue usuellement en s'assurant qu'aucune composante des forces (dans chaque direction de l'espace) n'excède une valeur seuil; dès lors que cette composante maximale, nommée F_{max} dans la figure, est inférieure à une valeur seuil fixée par l'utilisateur, on arrête l'optimisation. L'optimisation présentée dans la figure emploie l'algorithme du gradient conjugué [110] (mot-clé *Driver* = *Conjugate-Gradient* dans les fichiers d'entrée de *DFTB+*). Dans le cas des splines, on fixe la valeur de $F_{max} = 1 \cdot 10^{-4}$ unités atomiques, soit environ 0.005 eV/Å (mot-clé *MaxForceComponent*). Comme la convergence est plus ardue pour *NN-E*_{rep}, on double la valeur de ce seuil ($F_{max} = 2 \cdot 10^{-4}$ a.u. $\approx 0.01 \text{ eV/Å}$); on diminue également l'amplitude des modifications appliquées par l'algorithme (mot-clé *MaxAtomStep*= 0.0005) pour faciliter la convergence, au prix d'une optimisation plus longue.



FIGURE 8.9 – Évolution des énergies et des forces pour l'optimisation des géométries de deux nano-clusters Si_{14} .

Pour les splines, la figure confirme que l'optimisation de géométrie est très aisée, et les deux géométries convergent sans aucun problème après une bonne vingtaine de pas. Le constat est tout autre avec le potentiel répulsif de type réseau de neurones. Dans un cas $(Si_{14} n^{\circ}2)$, on rencontre tout simplement un plateau, où la force maximale vaut environ 0.02 eV/Å, que l'algorithme ne parvient pas à outrepasser avant le nombre maximal d'itérations. Dans l'autre cas $(Si_{14} n^{\circ}1)$, on parvient à converger après environ 800 itérations, mais de manière assez insatisfaisante. En effet, on transite par un premier minimum d'énergie autour de 300 pas, mais le système présente **au moins une** force interatomique trop élevée pour pouvoir considérer la géométrie comme stable. Après plusieurs autres centaines de pas, la composante maximale des forces F_{max} atteint enfin la valeur cible; les forces sont suffisamment petites, mais l'énergie totale n'est de toute évidence pas minimale. Aucune des situations rencontrées ne correspond donc à une géométrie stable.

Ces données appuient le constat effectué précédemment : le réseau de neurones NN- $E_{\rm rep}$ décrit plutôt mal le paysage énergétique des nano-clusters de Silicium pur. Celui-ci semble peu régulier, l'optimisation de géométrie ne permet pas systématiquement d'identifier des configurations d'énergie minimale.

En pratique, cette difficulté ne s'observe que pour les systèmes pour lesquels le réseau représente mal les forces; on ne le relève pas pour les cristaux, mais ce phénomène est très présent pour les nano-clusters. La mauvaise estimation des forces interatomiques semble indiquer que les énergies répulsives retournées varient relativement fortement en réponse à des changements minimes de configurations. Ainsi, on devine que la surface d'énergie potentielle du réseau de neurones est accidentée dans la région des nano-clusters, et nous n'avons pas de raison d'assumer qu'il en va différemment pour les autres objets isolés. En conséquence, et bien qu'une itération soit aussi coûteuse numériquement indépendamment du potentiel répulsif utilisé (voir annexe C.2), le processus est globalement plus long avec les réseaux de neurones et semble moins fiable qu'avec le potentiel de paire.

Remarque : les paramètres de la partie électronique de DFTB sont rigoureusement identiques dans le cas des splines et de nos outils $NN-E_{\rm rep}$. Il semble donc raisonnable d'affirmer que ce sont les forces répulsives $F_{\rm rep}^{\rm NN}$ qui sont responsables des erreurs observées. Il est aussi possible que ces forces surestimées dans le cas du potentiel répulsif numérique amènent la géométrie du système dans des régions du paysage d'énergie potentielle où les contributions électroniques à l'énergie DFTB ($E_{\rm BS}$ et $E_{\rm coul}$) sont en défaut et peinent également à converger. Que la réalité tienne de l'une ou l'autre de ces raisons, ou même d'une combinaison des deux, il s'agit d'un phénomène indétecté pour les splines : il est donc clair que nos outils sont en cause, d'une manière ou d'une autre.

8.2.2 Prédiction de la compacité des nano-clusters

Enfin, on peut également s'interroger sur les caractéristiques structurelles prédites par les différents outils investigués. Un bon candidat pour cela est le rayon de giration r_g des clusters les plus stables de chaque taille.

$$r_g^2 = \frac{1}{N^{\text{atomes}}} \sum_{i=1}^{N^{\text{atomes}}} \left(r_i - \bar{r}\right)^2.$$

Cette grandeur est directement liée à la compacité du nano-cluster, et présente dans le cas du Silicium une propriété intéressante : entre $N^{\text{atomes}} = 26$ et $N^{\text{atomes}} = 27$, ce rayon chute fortement et indique un réarrangement fondamental de géométrie [111]. Pour les petits clusters, beaucoup d'atomes sont en surface; à partir de $N^{\text{atomes}} = 27$, ce comportement devient instable et les clusters se reconfigurent vers une forme plus compacte au cœur quasi-cristallin. Ce phénomène s'accompagne d'un profond changement dans les propriétés optiques, géométriques et électroniques : le rayon de giration est donc un indicateur structurel pertinent.

On compare dans la figure 8.10 les R_g des minima de chaque taille de clusters optimisés respectivement avec *VASP*, les différents potentiels répulsifs *DFTB* (splines et *NN-E*_{rep}) et les potentiels purement paramétriques (*NNP* et *GAP*).



FIGURE 8.10 – Rayons de giration des nano-clusters de Silicium.

De toutes les approches représentées, les splines sont une nouvelle fois les plus fiables : exception faite d'une erreur sensible pour $N^{\text{atomes}} = 14$ et d'une transition légèrement amortie pour $N^{\text{atomes}} \in [26, 28]$, les rayons de giration prédits sont très globalement comparables à la référence VASP. Il est néanmoins intéressant de noter pour plus tard leur tendance à surestimer, légèrement mais de manière consistante, r_g pour une majorité de tailles de clusters.

Après le potentiel répulsif de paire, il semble que GAP soit le deuxième plus précis. Les seules erreurs significatives effectuées sont dans la région proche de la transition configurationnelle ($N^{\text{atomes}} \in [25, 28]$), et ne sont pas trop préoccupantes. Les fluctuations par rapport à la référence sont légèrement plus marquées que pour les splines, mais restent acceptables. On peut partiellement attribuer cette bonne représentation des propriétés structurelles à la présence de nombreuses surfaces au sein des données de référence utilisées pour optimiser les paramètres de ce potentiel.

Les deux potentiels employant des réseaux de neurones (NNP et $NN-E_{rep}$) sont de loin les moins robustes, même si NNP semble être le seul potentiel parvenant à rendre compte de la transition à $N^{\text{atomes}} = 27$. On observe avec ces deux approches des fluctuations importantes, plus marquées avec notre outil mais bien plus fréquentes pour NNP^5 . En particulier, le potentiel répulsif numérique $NN-E_{rep}$ effectue une erreur très importante pour $N^{\text{atomes}} = 30$, qui s'accompagne de quelques erreurs moins marquées mais aussi de nombreux accords parfaits avec VASP. Notre approche semble donc avoir extrait quelques caractéristiques des nano-clusters de la base de données d'apprentissage, mais de manière trop variable pour que l'on puisse la considérer comme fiable ou même robuste.

Portons désormais notre attention spécifiquement sur les rayons obtenus avec les potentiels répulsifs (splines et $NN-E_{rep}$), et plus particulièrement sur le cas $N^{\text{atomes}} = 30$. L'erreur marquée commise par le potentiel répulsif de type réseau de neurones est par ailleurs consistante sur l'intégralité des potentiels répulsifs de type réseaux de neurones entraînés avec le même jeu de référence; cette lacune semble intrinsèquement liée au jeu de données d'apprentissage. On présente dans la figure 8.11 l'énergie répulsive par atome pour deux configurations de Si₃₀ : une première, compacte, que VASP et les splines prédisent comme la plus stable, et une autre structure allongée à laquelle le potentiel $NN-E_{rep}$ attribue ce qualificatif.

On remarque d'une part que le réseau de neurones $NN-E_{\rm rep}$ prédit des énergies répulsives globalement plus élevées que le potentiel de paire. Les énergies répulsives de référence utilisées pour entraîner le réseau de neurones sont d'une amplitude similaire -8.56 eV pour le nano-cluster compact et 12.83 eV pour l'autre. Cet ordre de grandeur n'est donc pas un problème *en soi*, mais les valeurs exactes retournées par le réseau de neurones sont erronées.

Pour comprendre ce phénomène plus en détail, portons notre intérêt sur l'énergie répulsive par atome prédite par les deux potentiels répulsifs étudiés. Pour le nano-cluster compact, le réseau attribue (Fig. 8.11a) une forte énergie répulsive aux atomes placés en surface et retourne des valeurs bien plus modérées pour les atomes de cœur, dont l'environnement atomique est similaire à celui d'une phase condensée. En raison de leur forme de paire, les splines prédisent (Fig. 8.11b) une énergie répulsive relativement faible aux atomes de surface; cela s'effectue en éloignant les atomes du centre de masse du cluster, ce que confirme la figure 8.10 pour laquelle on remarque une tendance à sousévaluer légèrement la compacité des clusters.

La surestimation par le réseau $NN-E_{\rm rep}$ de l'énergie répulsive des atomes placés en surface se poursuit pour le nano-cluster Si₃₀ allongé (Fig. 8.11c); ceux-ci ci sont en revanche moins fréquents que pour le nano-cluster compact, et l'énergie répulsive globale prédite est en conséquence plus faible. Les atomes dont l'environnement semble quasiment cristallin - principalement autour du centre du cluster - voient leurs contributions à l'énergie répulsive être relativement faibles. Là encore, le phénomène inverse est observé

^{5.} Dans le cas du potentiel NNP, la présence de quelques surfaces au sein du jeu de données d'apprentissage ne semble pas suffisante pour permettre une bonne représentation des propriétés structurelles (à l'inverse de GAP).



 E_{rep} [ev/atome]

(a) Prédiction du réseau de neurones $NN-E_{\rm rep}$ pour un nano-cluster Si₃₀ compact, d'énergie répulsive totale $E_{\rm rep}^{\rm tot} = 9.103$ eV.



(c) Prédiction du réseau de neurones $NN-E_{\rm rep}$ pour un nano-cluster Si₃₀ allongé, d'énergie répulsive totale $E_{\rm rep}^{\rm tot} = 7.841$ eV.



0.002 0.004 0.006 0.008 0.010 0.012 0.014 0.016 *E_{rep}* [ev/atome]

(b) Prédiction des splines pour un nanocluster Si₃₀ compact, d'énergie répulsive totale $E_{\rm rep}^{\rm tot} = 0.180$ eV.



(d) Prédiction des splines pour un nanocluster Si₃₀ allongé, d'énergie répulsive totale $E_{\rm rep}^{\rm tot} = 0.098$ eV.

FIGURE 8.11 - Énergies répulsives par atome calculées pour plusieurs nano-clusters Si₃₀.

avec les splines (Fig. 8.11d), qui attribuent des contributions importantes à ces mêmes atomes alors que l'impact de ceux placés en surface sur l'énergie répulsive est quasiment nul.

Pour résumer, l'erreur forte observée dans la figure 8.10 pour le rayon de giration du cluster Si_{30} de plus basse énergie n'est pas liée à une explosion du nano-cluster, mais bien au choix d'un mauvais candidat par le réseau de neurones NN- E_{rep} . Celui-ci privilégie fortement les configurations présentant des environnements condensés, dans le sens où un plus grand nombre d'atome est partiellement entouré dans des directions plus variées de l'espace; nous sommes ici face à un cas curieux ou un plus grand nombre d'environnement condensés n'est pas directement lié à une plus grande compacité. Il semble en revanche que cette tendance soit indépendante de la coordinence effective des atomes, et le potentiel numérique montre une préférence pour les atomes entourés dans chaque direction de l'espace; on peut supposer une possible importance des contributions angulaires à la description des environnements atomiques.

La représentation des autres tailles de nano-clusters est plus satisfaisante : on note çà et là des fluctuations légèrement trop fortes rendant le potentiel de paire plus fiable, mais

une majorité des rayons prédits par le potentiel répulsif de type réseau de neurones sont très proches des valeurs de référence, et sont même fréquemment identiques. Le rayon d'inflexion à $N^{\text{atomes}} = 27$ n'est pas mieux reproduit que pour les splines; il est également amorti entre $N^{\text{atomes}} = 26$ et $N^{\text{atomes}} = 28$. Globalement, le réseau tend à privilégier les configurations pour lesquelles un maximum d'atomes ont un environnement proche de ceux observés dans les phases condensées. Ainsi, lorsque la taille des clusters augmente, et que l'importance des effets de surface s'amoindrit au profit d'une structure de cœur cristalline, le réseau devient plus fiable et semble même souvent surclasser le potentiel de paire.

D'un point de vue plus global, la représentation des systèmes non-périodiques semble hors de portée des trois approches par apprentissage automatique étudiées dans cette section (DFTB+NN- E_{rep} , NNP ou GAP). Ces outils se basent sur une description de l'environnement atomique local très efficaces pour les systèmes périodiques, mais montrant des preuves d'insuffisances marquées pour les systèmes isolés : mauvaise discrimination énergétique (figures 8.7 et 8.8) pour NNP et GAP, inefficacité numérique dans l'optimisation de géométries (figure 8.9) pour les trois potentiels ⁶ ou encore prédiction structurelle fragile (figure 8.10) pour NNP et NN- E_{rep} .

On pourrait espérer combler partiellement cette lacune en s'orientant vers une description de l'environnement atomiques plus adaptée (au risque de réduire notre précision pour les systèmes condensés), ou encore en introduisant plus de surfaces au sein du jeu de données d'apprentissage de notre outil : dans les cas de *NNP* et *GAP*, l'entraînement sur ce genre de système semble aider à améliorer les rayons de giration prédits. L'apparition de surfaces s'accompagne de comportements très variés (un atome de cœur est très différent d'un atome de surface), et notre construction ne nous permet de toute évidence pas d'en tenir complètement compte. Dans l'idéal, des données énergétiques à l'échelle de chaque atome - pour l'instant inaccessibles - pourraient aider cette discrimination, là où l'utilisation conjointe d'énergies répulsives de boites (informations non locales) et de forces atomiques répulsives (locales) est manifestement insuffisante.

Au final, le potentiel répulsif de type splines reste donc très largement préférable à tous les autres pour sa reproduction très acceptable des propriétés énergétiques et structurelles des nano-clusters étudiés ici. Au-delà de ces résultats spécifiques, c'est surtout la robustesse et la fiabilité des splines, convergeant **toujours** rapidement vers un minimum d'énergie, qui leur offre un rôle de choix pour les applications concentrées sur des systèmes isolés au sens large.

8.3 Systèmes désordonnés de Silicium pur

Étudions enfin la capacité de nos outils à rendre compte du comportement de phases désordonnées du Silicium pur. On s'intéresse pour cela aux liquides mais également à

^{6.} L'efficacité numérique et la difficulté à converger les géométries de clusters n'ont été présentées que pour notre approche NN- E_{rep} , car il s'agit du point focal des présents travaux, mais on observe des difficultés très analogues - et parfois même plus marquées - pour NNP et GAP.

quelques amorphes métastables⁷.

8.3.1 Silicium liquide

Portons dans un premier temps notre attention sur la représentation des liquides. La distribution des distances au sein du système est en général une grandeur pertinente : plus la température augmente, plus les atomes sont mobiles et un désordre caractéristique apparaît. Dès lors, on s'assure (fig. 8.12) que les différents potentiels répulsifs reproduisent correctement le désordre prédit par VASP. Le liquide étudié est le même dans chacun des cas : 128 atomes dans la boîte de simulation, obtenue à partir d'une configuration initiale de Silicium diamant chauffée à T = 3000 K pendant 30 ps avec VASP. À partir de cette même configuration initiale, on effectue avec toutes les méthodes étudiées (VASP, splines, $NN-E_{\rm rep}$, NNP et GAP) une dynamique moléculaire pour quatre températures différentes : T = 2000 K, T = 2500 K, T = 3000 K et T = 3500 K. On choisit de s'intéresser à des températures jusqu'à T = 3500 K pour étudier le comportement sur des plages de températures inconnues de notre outil $NN-E_{\rm rep}$, dont la température maximale présente au sein du jeu d'apprentissage est T = 3000 K.

Chacune de ces dynamiques moléculaires s'effectue sur une durée variant en fonction de la température; cela est dû à quelques difficultés rencontrées avec DFTB+, pour lequel il arrive que les calculs auto-consistants ne convergent pas pour des systèmes trop désordonnés. On extrait de ces dynamiques moléculaires la fonction de distribution radiale g(r). On n'observe pour aucun potentiel un rapprochement excessif d'atomes (g(r) = 0aux courtes distances); par souci de lisibilité, on démarre les tracés autour de $r \approx 2$ Å. On reporte également, à titre indicatif, la fonction de distribution radiale obtenue expérimentalement [112] pour du Silicium pur liquide à T = 1733 K.

Pour chaque température, on note d'une part que les calculs présentés, indépendamment de la méthode choisie, prédisent globalement une distance aux premiers voisins, $r \approx 2.50$ Å pour lequel g(r) est maximal, très similaire aux données expérimentales. Cette distance est d'ailleurs consistante sur l'intégralité de la plage de température étudiée. Il est aussi intéressant de noter que la distance aux premiers voisins est très proche du rayon de coupure des splines, $r_{\rm cut}^{\rm splines} = 2.54$ Å; l'impact du potentiel répulsif de paire sur les distributions observées est sans doute minime.

Il est aussi apparent que VASP sous-estime l'ordre global au sein du liquide par rapport à l'expérience (pics moins marqués) - ce qui paraît cohérent à T = 2000 K au vu de l'écart significatif entre la température expérimentale et celles simulées. On observe en revanche que *DFTB*, indépendamment du potentiel répulsif (splines et $NN-E_{rep}$), surestime légèrement l'ordre du système. On peut raisonnablement supposer que l'ordre aux longues distances (au-delà de 3 Å) est dominé par les contributions électroniques à l'énergie totale *DFTB*; nos outils ne semblent pas en mesure de corriger complètement ceci, même s'ils semblent atténuer légèrement l'ordre prédit à $T \leq 3000$ K par rapport aux splines.

^{7.} Systèmes dans un minimum d'énergie local, d'apparence stable mais pouvant aisément se reconfigurer en cas de perturbation.



FIGURE 8.12 – Fonctions de distribution radiale du Silicium liquide à différentes températures.

Si on compare les valeurs obtenues avec VASP, le potentiel répulsif numérique $NN-E_{\rm rep}$ fonctionne de manière satisfaisante pour les températures inférieures ou égales à la température du jeu de données de référence ($T \leq 3000$ K, fig. 8.12a - 8.12c). Il semble même surclasser imperceptiblement les splines sur cette plage de températures, en ordonnant légèrement moins les liquides (pics moins marqués); l'écart entre les deux potentiels répulsifs est néanmoins trop faible pour être significatif. En revanche, pour une température supérieure aux données de référence (T = 3500 K, 8.12d), on remarque que le réseau tend à sur-ordonner le liquide : le pic de premiers voisins est sensiblement plus haut et moins large que la référence, et on observe des fluctuations non-physiques notamment autour du rayon de coupure $r_{cut} = 3$ Å.

De manière générale, plus la température est élevée et plus la probabilité d'explorer des régions variées de l'espace des configurations est importante (voir 7.1.2, eq. (7.3)). Plus simplement, il est probable en simulant un système suffisamment longtemps à une température donnée de voir apparaître localement toutes les configurations caractéristiques des températures inférieures. C'est pour cela que l'on génère usuellement des données de référence à assez haute température lorsque l'on veut optimiser un outil d'apprentissage automatique. En revanche, la transférabilité aux températures plus élevées est incertaine : ceci semble se confirmer partiellement ici. Le problème le plus gênant lorsque l'on extrapole à des températures inconnues est habituellement d'obtenir des potentiels inaptes à représenter efficacement les courtes distances et présentant pour celles-ci des pics non-physiques. Dans notre cas, ce comportement semble convenablement prévenu par l'inclusion de systèmes comprenant de très faibles distances (courbes énergie vs. volume, paire d'atomes) dans le jeu de données d'apprentissage⁸.

Les approches purement numériques (NNP et GAP) sont quant à elles quasiment identiques à la référence VASP pour l'intégralité de la plage de températures étudiées. NNP paraît légèrement plus précis que GAP pour $T \leq 2500$ K, et la tendance s'inverse de peu pour les plus hautes températures. Ici, on tire pleinement parti de la flexibilité des méthodes d'apprentissage automatique en jeu, puisqu'elles sont complètement affranchies des limitations apportées par les contributions électroniques observées dans le cas de DFTB. La température maximale des données de référence de NNP (le double de la température de fusion du Silicium, soit ~ 3375 K) est assez proche de T = 3500 K, et la bonne distribution observée pour ce potentiel n'est pas surprenante. Le cas de GAP est plus ardu à interpréter. Le potentiel a été testé, d'après [74], seulement à une température avoisinant T = 2000 K. La température exacte du jeu d'apprentissage est inconnue, mais est probablement bien supérieure à T = 2000 K au vu du très bon accord avec VASP jusqu'à T = 3500 K.

Par construction, les réseaux de neurones sont à même de tenir compte des angles formés par les triplets d'atomes. Afin d'être péremptoire quand aux craintes légitimes que cela soulève, il est important de s'assurer que nos outils ne privilégient pas d'angles particuliers. On présente donc (fig. 8.13) la distribution angulaire ⁹ de liquides à différentes températures obtenues avec chacune des méthodes (*VASP*, splines, *NN-E*_{rep}, *NNP* et *GAP*).

Ces fonctions de distribution angulaire sont issues des mêmes dynamiques moléculaires que les fonctions de distribution radiale (Fig. 8.12), et l'on réduit simplement le rayon de coupure à 3 Å afin de ne considérer que les angles formés avec les premiers voisins.

Pour les températures les plus faibles ($T \leq 2500$ K, fig. 8.13a 8.13b) on remarque que l'approche employant le réseau de neurones NN- $E_{\rm rep}$ surestime l'importance des angles au-delà de $\pi/2$ par rapport à la référence VASP. Comme on observe également ce même comportement en plus marqué pour les splines, la prise en compte explicite des angles pour le réseau ne semble pas en cause. Les distances prises en compte pour déterminer la distribution angulaire sont plutôt faibles ($r_{\rm cut} = 3$ Å) : il n'est pas évident que ces gammes de distances soient dominées par les contributions électroniques à l'énergie *DFTB*. Même en admettant que ce soit effectivement le cas, le potentiel répulsif numérique devrait être

^{8.} Pour les calculs présentés dans ce chapitre, nous avons constamment vérifié que le potentiel *ad-hoc* de courte portée (voir 7.1.5) n'intervenait pas; les résultats observés sont donc liés uniquement à la prédiction de $NN-E_{rep}$.

^{9.} Équivalent de la fonction de distribution radiale, où l'on se concentre sur les angles plutôt que les distances.



FIGURE 8.13 – Fonctions de distribution angulaire du Silicium liquide à différentes températures.

en principe capable de corriger les erreurs à une portée aussi réduite. La seule correction significative apportée est autour de $\pi/4$, angles assez typiques des atomes sur-coordinés et pour lesquels la très forte sous-estimation de la distribution angulaire des splines est corrigée par nos outils.

Cette amélioration est d'autant plus visible pour les températures plus élevées ($T \geq 3000$ K, fig. 8.13c 8.13d). Dans ce domaine de température, notre outil reproduit assez fidèlement la distribution angulaire de référence, mieux que ne le fait le potentiel de type splines. Ici, la transférabilité correcte aux températures élevées (T = 3500 K) est surprenante : les angles tendent évidemment à s'uniformiser à mesure que la température - et donc le désordre - croît, et NN- $E_{\rm rep}$ semble saisir et retranscrire ce phénomène. Cette caractéristique s'avère relativement ardue à expliquer car toutes les données incluant des courtes distances typiques des hautes températures (énergies vs. volume et paire d'atomes) n'incluent aucune variabilité angulaire. En conséquence, rien ne garantit dans notre méthodologie de correctement reproduire les angles aux températures supérieures à

T = 3000 K. Ainsi, toute approche similaire appliquée à d'autres éléments chimiques ou systèmes requiert une attention particulière dans le cas, assez marginal, où les prédictions du potentiel aux très hautes températures sont capitales.

Par ailleurs, on note là encore de très bons résultats pour les potentiels purement paramétriques (*NNP* et *GAP*). Tout comme dans le cas des fonctions de distribution radiale, *NNP* semble légèrement plus précis que *GAP* à $T \leq 2500K$, et cet écart s'annule à mesure qu'on augmente la température.

Globalement, le constat est le même pour les distributions radiales et angulaires : d'une part, aucune méthode investiguée ne produit de comportement aberrant ou non physique. Il apparaît également que les outils d'apprentissage automatique ($NN-E_{rep}$, NNP et GAP) sont plus à même de rendre compte des propriétés des liquides, grâce à leur capacité à inclure facilement des systèmes désordonnés dans leurs jeux d'apprentissages respectifs. La précision est néanmoins moins marquée pour nos outils $NN-E_{rep}$ que pour les deux autres approches purement paramétriques, car nous nous limitons aux interactions à courte portée et sommes tributaires des limitations des contributions électroniques à l'énergie DFTB.

8.3.2 Configurations amorphes

Présentons enfin les tests effectués sur les systèmes amorphes. Soulignons qu'il n'existe pas **un** silicium amorphe de référence. Ces systèmes sont désordonnés, métastables, et leurs propriétés, en commençant par leur structure, peuvent dépendre de la technique utilisée pour les produire (dépôt sur une surface, irradiation laser, par implantation ionique dans un cristal, ...).

Néanmoins, la plupart des échantillons sont caractérisés par les propriétés suivantes : une densité inférieure à celle du cristal (de l'ordre de quelques % par rapport à la densité expérimentale du Silicium diamant, 2.33 g.cm⁻³), une coordinence inférieure à 4 [113,114], un angle Si-Si-Si autour de 108° avec une déviation typique de ~ 10°(voir par exemple [115]).

Les systèmes amorphes utilisés lors des présents travaux ont été construits ainsi : nous disposions de configurations obtenues après trempe rapide (chute de température de 10^{11} K.s⁻¹) d'un liquide à haute température (T = 3500 K) modélisé par un potentiel de Tersoff [116] pour des très grands systèmes (volume d'environ 9 nm³). Ces configurations après trempe sont évidemment trop grandes pour être étudiées avec une approche premier principe type VASP, et l'on en extrait des échantillons plus petits contenant approximativement une centaine d'atomes. On applique des conditions aux limites périodiques à chaque échantillon en éliminant les atomes trop proches (distances inférieures à 2 Å). On relaxe ensuite les positions atomiques en effectuant des dynamiques moléculaires classiques à basse température (T = 300 K) pendant 10 ps. On obtient enfin une structure stable en minimisant la cellule et les positions atomiques avec VASP; après vérification finale, les pressions observées sont de l'ordre de quelques MPa.

En suivant cette procédure, on ne s'attend pas à reproduire précisément les carac-

téristiques du Silicium amorphe. Néanmoins, on dispose de plusieurs échantillons d'un amorphe en volume (*«bulk silicon»*) contenant des défauts avec des coordinences typiques allant de 3 à 6. L'introduction de ces systèmes dans la base de donnée permet de diversifier les environnements atomiques en incluant des configurations avec défauts.

Les structures des systèmes amorphes ainsi générés ont été ré-optimisées (en gardant cette fois un volume constant) avec les différentes techniques en discussion. Les résultats pour les trois systèmes amorphes contenant respectivement 93, 96 et 100 atomes (présents dans les données d'apprentissage de notre potentiel répulsif NN-Erep), ainsi que les résultats pour un autre système amorphe avec 102 atomes (non inclus au jeu d'apprentissage de NN-Erep) sont rassemblés dans la table 8.1.

Taille	VASP	Splines	$NN-E_{rep}$	NNP	GAP	Densité
						$[g.cm^{-3}]$
93 atomes	4.18 voisins	4.10 voisins	4.10 voisins	4.03 voisins	4.08 voisins	2.31
Angle moyen	106.9°	107.3°	107.0°	107.9°	106.9°	
Disp. angulaire	18.1°	17.1°	18.0°	16.2°	17.9°	
96 atomes	3.99 voisins	3.97 voisins	3.97 voisins	3.95 voisins	3.97 voisins	2.29
Angle moyen	107.1°	107.8°	107.3°	107.9°	107.9°	
Disp. angulaire	16.6°	15.4°	15.9°	16.0°	15.0°	
100 atomes	3.96 voisins	3.96 voisins	3.96 voisins	3.95 voisins	3.92 voisins	2.28
Angle moyen	108.1°	108.0°	108.1°	108.6°	108.5°	
Disp. angulaire	15.8°	15.8°	15.2°	14.2°	14.6°	
102 atomes	4.00 voisins	3.96 voisins	4.02 voisins	4.02 voisins	4.04 voisins	2.34
Angle moyen	107.7°	108.1°	107.5°	107.0°	107.9°	
Disp. angulaire	16.7°	16.0°	17.3°	18.4°	16.3°	

TABLE 8.1 – Caractéristiques structurelles des amorphes relaxés par chaque méthode. Les coordinences et les distributions d'angles sont calculées en tenant compte des voisins jusqu'à 2.85 Å (position moyenne du premier minimum des distributions radiales).

On remarque des coordinences et des densités relativement proches de celles des systèmes expérimentaux, à part l'amorphe à 93 atomes sur-coordiné pour chaque méthode, et la dispersion autour de l'angle moyen est nettement trop grande dans nos calculs. Ceci traduit un excès de défauts dans nos échantillons par rapport à une structure compatible avec les données expérimentales. Nous obtenons néanmoins avec toutes les méthodes testées des structures **stables** proches de celles initialement fournies par le code VASP.

Les valeurs exactes prédites pour chacune de ces informations structurelles seraient trop longues à détailler exhaustivement, mais soulignons quelques points semblant intéressants. D'une part, notre approche NN- $E_{\rm rep}$ est très proche de la référence VASP pour toutes les propriétés étudiées, même pour l'amorphe à 102 atomes; ce résultat est très encourageant, car il indique deux choses. La bonne reproduction des systèmes connus prouve que notre outil est effectivement en mesure d'apprendre et restituer des caractéristiques précises pour des configurations désordonnées connues. La bonne capacité de prédiction pour un système proche mais néanmoins inconnu (102 atomes) permet aussi de supposer une relative capacité de généralisation. En somme, notre outil est apte à reproduire les propriétés structurelles de la méthode de référence, peut-être même imperceptiblement mieux que les autres outils étudiés. Du côté des approches purement numériques, on peut aussi noter prédictions structurelles généralement satisfaisantes; le fait que GAP ne soit pas plus précis que NNP est toutefois assez surprenant, car le premier contient des amorphes dans son jeu de données de référence, mais pas le second.

On s'intéresse finalement à la densité d'états vibrationnels de ces mêmes amorphes. Pour ces systèmes désordonnés métastables, les propriétés vibrationnelles sont accessibles expérimentalement [117]; leur caractérisation par nos outils est donc importante, et fournit un point de comparaison intéressant. En partant des géométries optimisées, nous trouvons que dans chaque cas une cellule $2 \times 1 \times 1$ est suffisante pour décrire les vibrations. On a donc des tailles de systèmes respectivement égales à 186, 192, 200 et 204 atomes. On applique avec *phonopy* (via l'interface *phonoLAMMPS* pour *NNP* et *GAP*) des déplacements individuels d'atomes qui génèrent des forces permettant de calculer une courbe de dispersion des phonons en plusieurs points de la première zone de Brillouin (là encore avec une grille *Monkhorst-Pack* de $30 \times 30 \times 30$ points dans l'espace réciproque) qu'on intègre sur les fréquences pour obtenir la densité d'états vibrationnels.

On présente les résultats obtenus avec le potentiel répulsif de paire, le potentiel répulsif de type réseau de neurones et VASP dans la figure 8.14. Les densités d'états vibrationnels obtenus avec NNP et GAP sont fournis en annexe E.3, afin de garder des figures intelligibles.

Pour des systèmes aussi désordonnés que le Silicium amorphe, les densités d'états vibrationnels sont difficiles à lire en raison des fortes fluctuations. Globalement, on parvient néanmoins à observer que les deux potentiels répulsifs (splines et $NN-E_{\rm rep}$) font preuve de comportements très similaires. On note bien, çà et là, quelques divergences; sans surprise, celles-ci concernent peu les modes acoustiques. Nous avons vu pour les monocristaux (Fig. 8.4) que ceux-ci sont régis majoritairement les contributions électroniques à l'énergie DFTB, et ils semblent dans le cas des amorphes bien représentés par ces derniers. En conséquence, le choix du potentiel répulsif impacte là aussi assez peu les modes où $\omega < 7.5$ THz. Pour ces faibles fréquences, les résultats expérimentaux [117] prédisent un premier maximum autour de $\omega \approx 20$ meV ≈ 4.8 Thz. VASP reproduit bien ce comportement, et les potentiels répulsifs DFTB montrent eux aussi un accord raisonnable avec cette valeur.

En portant notre attention sur modes optiques, on parvient à remarquer quelques nuances plus notables entre les deux potentiels répulsifs. Ainsi, les amplitudes et fréquences peuvent sensiblement varier, mais ces fluctuations sont assez minimes et penchent tantôt en faveur des splines, tantôt en faveur de $NN-E_{\rm rep}$. La forme globale de chaque densité d'états vibrationnels étudiée est donc satisfaisante pour ces deux outils, qui paraissent au final assez équivalents pour cette propriété. Expérimentalement, on prédit un second maximum pour les densités d'états vibrationnels pour $\omega \approx 60 \text{ meV} \approx 14.5 \text{ Thz}$; ici encore VASP montre un accord correct avec cette valeur. Les différents potentiels répulsifs montrent eux aussi des maxima autour de cette fréquence : les splines semblent les plus proches de cette valeur pour les quatre amorphes, alors que le réseau de neurones $NN-E_{\rm rep}$ y parvient parfaitement pour 96 atomes seulement.

Le constat effectué est assez encourageant; la représentation des densités d'états vi-



FIGURE 8.14 – Densité d'états vibrationnels de plusieurs systèmes amorphes de Silicium pour différents potentiels répulsifs *DFTB*.

brationnelles n'était pas un objectif *per se* lors du processus d'entraînement, mais découle des forces interatomiques. Les résultats observés indiquent que ces dernières semblent bien reproduites par nos outils, et on ne note pas d'erreur rédhibitoire. Cette assertion se vérifie aussi dans le cas de l'amorphe à 102 atomes, absent des données de référence de $NN-E_{\rm rep}$. Notre potentiel répulsif ne fait donc pas *«que»* reproduire fidèlement les forces répulsives d'amorphes connus, mais semble bien être en mesure de représenter d'autres configurations conceptuellement proches mais localement différentes.

Les approches purement paramétriques, présentées en annexe E.3, sont toutes deux relativement analogues et reproduisent assez fidèlement les densités d'états vibrationnels des amorphes. En particulier, NNP reproduit très précisément les modes acoustiques; ce comportement est assez étonnant car aucun amorphe ne figure parmi ses données de référence. Ce manque se fait sommairement sentir pour le système contenant 102 atome, pour lequel on observe des fréquences négatives mais aussi et surtout de nombreuses fréquences résiduelles très élevées ($\omega \in [20, 105]$ THz) que nous avons partiellement coupées pour garder une figure lisible. Cela indique une instabilité des forces interatomiques, et donc que la géométrie n'est pas suffisamment relaxée. Pour *GAP*, ce comportement n'apparaît pas et les résultats sont donc généralement satisfaisants. Enfin, ces deux méthodes reproduisent bien le premier maximum expérimental [117] ($\omega \approx 4.8$ THz) mais sous-estiment le second ($\omega \approx 14.5$ THz).

Au global, la prédiction des densités d'états vibrationnels pour les amorphes est acceptable dans chacun des cas étudiés. Les méthodes DFTB (splines et $NN-E_{rep}$) sont extrêmement similaires, et les approches purement numériques (NNP et GAP) montrent des comportements sensiblement plus variés mais satisfaisants.

Avant de conclure, résumons la capacité de reproduction observée pour chacun des outils sur les systèmes désordonnés. Pour les liquides (fig. 8.12 et 8.13), les approches purement paramétriques (NNP et GAP) jouissent d'une plus grande flexibilité et sont les plus performantes. Des deux approches DFTB, le potentiel numérique répulsif $NN-E_{\rm rep}$ montre une précision sensiblement accrue par rapport aux splines, mais une attention particulière est requise aux très hautes températures, lorsque l'on entre en régime extrapolatoire. Une bonne partie des caractéristiques des liquides sont régies par les contributions électroniques de DFTB; la correction apportée par notre approche est sensible par rapport aux splines, mais fait pâle figure au regard de NNP et GAP.

Pour les amorphes, l'étude des densités d'états vibrationnels est assez comparable pour les quatre méthodes qui reproduisent bien la référence VASP; cela ne nous permet pas de désigner un potentiel clairement plus performant. L'étude des facteurs structurels nous permet en revanche de clairement établir l'importance du jeu d'apprentissage pour la prédiction de la structure locale : en effet, NN- E_{rep} performe de manière très satisfaisante. Les approches purement paramétriques (NNP et GAP) sont très comparables entre elles et incluent des systèmes désordonnés dans leur jeux d'apprentissage. Au global, toutes les méthodes par apprentissage automatique testées rendent donc mieux compte de la structure locale. Ceci met en avant l'intérêt de pouvoir inclure des systèmes désordonnés lorsque l'on optimise un potentiel, qu'il soit répulsif ou non.

Chapitre 9

Conclusion

Les travaux de cette thèse contribuent à créer un cadre méthodologique afin d'améliorer un terme du formalisme *Density-Functional Tight-Binding*, le potentiel répulsif, en permettant une représentation à plusieurs corps employant des outils d'apprentissage automatique en alternative aux définitions existantes sous forme d'interactions de paires.

Une démarche globale est introduite pour les obtenir, qui peut s'appliquer à tous les ordres de développement de la méthode. On en présente une application pour le développement au second ordre (*SCC-DFTB*), et un nouveau potentiel répulsif de portée générale pour le *Silicium* a été développé et étudié. Cette méthodologie fait usage des fonctions de symétrie de Behler-Parrinello greffées à des réseaux de neurones de haute dimension, et permet par sa souplesse d'inclure une grande complexité au sein des données employées pour optimiser ses potentiels répulsifs. Elle autorise également à l'utilisateur de choisir le degré d'intuition physique qu'il souhaite intégrer - via le choix des paramètres des descripteurs - tout en garantissant de pouvoir se reposer sur des processus de génération et de sélection systématiques si nécessaire. L'emploi de réseaux de neurones requiert quant à lui une certaine expertise lors de la construction d'un nouveau potentiel répulsif mais assure dans le même temps un coût numérique réduit et une portabilité maximale.

Nous situons tout d'abord les nombreux outils conceptuels intervenant dans ces travaux les uns par rapport aux autres, puis nous présentons leur utilisation conjointe. Les détails de celle-ci, la complexité sous-jacente et les écueils inhérents à l'application des réseaux de neurones de haute dimension aux potentiels répulsifs sont discutés extensivement. La complexité supplémentaire et la robustesse apportés par le formalisme DFTBpar rapport aux potentiels numériques utilisés usuellement en simulation atomistique classique sont notables et souvent non-triviales. Il est nécessaire d'adapter les routines en conséquence, et un cheminement clair est détaillé avec une emphase forte sur la logique sous-jacente : choix des données inclues, construction et sélection de descripteurs, de réseaux de neurones, algorithmes d'optimisation et évaluation de performance. Le code DFTB+ est modifié de manière à supporter l'utilisation de potentiels répulsifs numériques ; les modifications sont minimales, au sein du code comme du côté utilisateur, et la portabilité et la performance numérique se montrent très satisfaisants.

En pratique, ces travaux se sont divisés en plusieurs étapes : un effort notable a

dès le départ été consacré au développement d'outils numériques sur mesure à même de manipuler les concepts clés de notre approche. La construction d'une librairie évolutive centrée autour des notions de réseaux de neurones de haute dimension et de fonctions de symétrie centrées sur les atomes a été une étape importante, majoritairement dominée par la mise en place de l'apprentissage sur les forces répulsives (partie du code au final inutilisée, au profit du filtre de Kalman étendu).

En parallèle de cette tâche en ont été menées deux autres. D'une part, nous avons travaillé pour l'obtention de données de référence par des calculs *DFT* et *DFTB* pour le Silicium pur, graduellement enrichies au fil des mois. Par ailleurs, ces données ont été de tous temps employées pour entraîner des réseaux de neurones à plusieurs fins : guider l'obtention de nouvelles données, valider le bon fonctionnement de la librairie, évaluer la précision de chaque outil ou encore caractériser l'importance effective de chaque concept ou paramètre. Ce dernier point en particulier est très enrichissant, car l'emploi de méthodes d'apprentissage automatique s'accompagne presque toujours d'outils opaques, à l'utilité conceptuelle définie mais dont les nuances sont en pratique subtiles ; s'approprier chacun de ces concepts, de leurs fonctionnements internes à leurs importances relatives s'est avéré très exigeant.

Deux jalons symboliques ont marqué des tournants importants des présents travaux. Tout d'abord, la délégation des l'optimisation des réseaux de neurones à N2P2 a permis d'obtenir des potentiels répulsifs réellement utilisables. En conséquence, chaque nouvel apprentissage apportait un peu plus d'intuition scientifique que le précédent, et les préoccupations numériques ont peu à peu laissé place à des interrogations physiques. Ceci s'est finalement concrétisé avec l'interfaçage de la librairie au code DFTB+, à la suite duquel nous avons pu convertir ce travail de fond en résultats concrets.

Nous avons pu éprouver extensivement nos outils sur trois grandes classes de configurations (cristaux, nano-clusters et systèmes désordonnés). Pour chacune d'entre elles, nous avons identifié des propriétés pertinentes et confronté les performances à plusieurs autres approches : la méthode DFT de référence, les potentiels répulsifs de paire utilisés jusqu'alors dans DFTB, un potentiel purement numérique employant les mêmes outils conceptuels que les nôtres et un second potentiel numérique de portée générale aux performances notables pour le Silicium pur. Dès lors, nous sommes en mesure d'identifier clairement les avantages et limitations de nos travaux, de cerner en conséquence leur intérêt réel.

D'une manière très globale, on peut tout d'abord conclure que la correction apportée *in-situ* par les potentiels répulsifs numériques est souvent sensible. On discute plus en détail de ces apports dans les paragraphes suivants; il est néanmoins important de souligner que la capacité à corriger les erreurs liées aux contributions électroniques du formalisme *DFTB* - notamment aux longues portées - est souvent insuffisante. On observe en effet de nombreuses propriétés dominées par les termes électroniques, et pour lesquelles l'impact du potentiel répulsif est moindre : traitement des défauts sur-coordinés, modes acoustiques des densités d'états vibrationnels, ordre des liquides ou encore classement de stabilité relative des monocristaux.

Le potentiel répulsif numérique permet d'améliorer sensiblement la qualité des phases

cristallines par rapport aux splines. Nous avons observé une faculté à reproduire les systèmes hors-équilibre, avec des courbes énergies vs. volume très proches des données VASP, des propriétés vibrationnelles assez similaires aux valeurs DFT et une amélioration des énergies de formation de défauts ponctuels (en particulier pour les interstitiels). L'impact sur l'état d'équilibre n'est pas notable, avec des écarts d'énergie relatifs même très légèrement meilleurs que pour le potentiel de paire. Les énergies et forces répulsives retournées par le réseau de neurones pour les cristaux semblent donc robustes et satisfaisantes.

La représentation des phases désordonnées n'est pas en reste, et le potentiel répulsif de type réseau de neurones semble transférable pour ces systèmes. Les géométries amorphes semblent assez bien représentées et leur variété configurationnelle forte ne paraît pas problématique. On observe ainsi une précision acceptable quand aux densités d'états vibrationnels obtenues avec nos outils, très analogue à celle des *splines*. L'écart observé pour les amorphes est un peu plus notable lorsque l'on s'intéresse à la prédiction des indicateurs structurels tels que la coordinence, l'angle moyen et la dispersion angulaire : le réseau de neurones reproduit assez fidèlement les nuances observées avec la méthode de référence, mieux que le potentiel de paire. Ceci n'est pas une surprise, car ce dernier n'a pas été (et n'aurait pas pu être) optimisé sur des systèmes désordonnés.

La représentation des liquides pour une large gamme de températures est légèrement améliorée par le potentiel répulsif numérique : les distributions angulaires et radiales bien que dominées aux longues distances par les contributions électroniques à l'énergie DFTB - sont toujours mieux représentées qu'avec les splines par rapport à la référence dans la plage de températures connues du potentiel numérique répulsif ($T \leq 3000$ K). Des phénomènes sporadiques font état d'extrapolation aux températures supérieures, tels que des distributions radiales irrégulières, sans pour autant s'accompagner de comportement rédhibitoires tels que le rapprochement excessif d'atomes.

Nous avons enfin pu éprouver notre méthodologie en l'appliquant à des systèmes non-périodiques : les nano-clusters de Silicium pur. On note une amélioration relative de l'évaluation énergétiques de configurations données par rapport aux splines, qui s'accompagne néanmoins d'une instabilité forte lors de l'optimisation de géométries isolées. Ces deux observations conjointes laissent tout d'abord deviner que, dans le cas de calculs statiques, les contributions électroniques à l'énergie *DFTB* constituent un garde-fou suffisant pour permettre des performances notables. Dans le cas où l'on doit recourir aux forces interatomiques, ce filet de sécurité devient insuffisant et les erreurs commises semblent trouver leurs sources dans de multiples phénomènes. Les fluctuations sensibles des forces indiquent une surface d'énergie potentielle accidentée, et l'étude détaillée de quelques configurations montre clairement que nos outils ne sont pas à même de traiter efficacement des effets de surface. Au regard de ces informations, et quand bien même nos outils présentent quelques résultats satisfaisants sur les systèmes isolés, les splines sont dans ce cas préférables pour leur robustesse, leur fiabilité et leur faculté à toujours converger par construction.

Bien que ces travaux ne présentent ni amélioration extrême ni régression dramatique, deux tendances se dessinent alors clairement : d'une part, la prévalence sensible de notre approche sur les potentiels de paire pour les phases condensées étudiées, et d'autre part son manque de fiabilité pour les objets isolés, pour lesquels les splines semblent plus performantes. L'objectif d'un unique potentiel pour le Silicium pur n'est dès lors pas tenu; l'implémentation d'un nouveau formalisme plus général très adapté à être enrichi par de nouveaux descripteurs ou d'autre méthodes d'apprentissage laisse toutefois cette perspective ouverte. Toutefois, l'impératif d'un outil peu coûteux est lui respecté puisque le coût numérique n'excède pas sensiblement celui des splines. Enfin, le détail des données de référence importantes, des écueils inhérents à cette approche et la proposition d'une démarche systématique pour générer un potentiel répulsif à partir de celles-ci font la part belle à la transposition de ces travaux pour de nouveaux éléments chimiques.

Un autre facteur important, la charge de travail requise, est à prendre en compte si l'on aspire à dresser une comparaison équitable entre notre approche et les potentiels répulsifs de paire. Dans le cas de ces derniers, on a besoin d'un nombre relativement restreint de données de référence (typiquement quelques dizaines de configurations tout au plus) nécessitant une attention soutenue dans leur obtention avant de réaliser une interpolation assez standard. On se concentre alors sur un nombre restreint de cas simples, d'une part car leur forme de paire peu flexible prohibe leur capacité à généraliser à une large variété de références, et d'autre part car il est trop ardu en principe d'identifier l'énergie de référence d'**une** paire d'atomes pour des systèmes non isolés ou non-symétriques; on obtient néanmoins assez aisément un outil, manquant souvent de flexibilité pour généraliser parfaitement mais extrêmement robuste.

Pour générer un potentiel répulsif numérique, les contraintes sont assez différentes puisque l'on peut a priori utiliser n'importe quelle combinaison de systèmes comme référence ; la flexibilité de ces outils leur permet de supporter une complexité accrue. Cela s'accompagne tout d'abord d'une augmentation conséquente des calculs préalables à effectuer. Ce point est accentué par la pauvre performance en extrapolation des outils d'apprentissage automatique : si l'on veut être transférable à de nombreux systèmes, il est recommandé de les inclure au jeu de données de référence, et la charge de travail augmente d'autant. Le choix des données de référence demande d'ailleurs une expertise assez spécifique à DFTB; en s'intéressant aux interactions répulsives de courte portée, il est intéressant d'inclure certaines données qui ne seraient pas nécessaires pour entraîner une approche purement numérique, comme des systèmes très compressés (par exemple courbes énergies vs. volume). Dans le même temps, et une fois les données de référence à disposition, un autre type d'expertise est requis pour l'optimisation des hyperparamètres de la méthode (fonctions de symétrie, structure du réseau, algorithme et paramètres d'optimisation). Des outils conceptuels, des automatisations et d'autres bonnes pratiques existent pour faciliter ces tâches; encore faut-il les connaître, et leur assimilation représente encore une charge de travail supplémentaire pour un néophyte. Pour résumer, les potentiels répulsifs numériques demandent plus de travail que leur homologues de paire, permettent en théorie une précision plus grande pour une plus grande gamme de systèmes au prix de l'abandon de la robustesse des potentiels à deux corps.

Au travers de la comparaison avec quelques résultats obtenus avec un réseau de neurones de Behler-Parrinello et d'un potentiel de type méthode à noyaux, tous deux optimisés pour reproduire l'énergie totale et les forces DFT du Silicium, nous avons pu d'autre part discuter des singularités de notre approche par rapport à des outils purement numériques. Nos travaux se concentrent sur le terme de l'énergie répulsive, qui joue en pratique le rôle de terme correctif au sein du formalise DFTB. Sans surprise, ces différences sémantiques fortes s'accompagnent de dissimilarités notoires. Ainsi, notre restriction à une contribution à l'énergie montre fréquemment des limitations à corriger effectivement des propriétés dominées par les termes électroniques - tels que la structure des liquides alors que les outils purement paramétriques jouissent pour leur part d'une flexibilité suffisante pour les mimer parfaitement. Dans le même temps, il est fréquent de rencontrer des situations pour lesquelles les approximations des termes électroniques de l'énergie DFTB sont robustes, auquel cas le travail à accomplir par nos outils est moindre. Ainsi, ces mêmes termes électroniques assistent notre approche en reposant sur une intuition physique forte, assurent souvent la prise en compte de la majorité de la complexité et préviennent donc l'apparition de comportements aberrants, même dans certains régimes extrapolatoires. Les approches purement paramétriques ne profitent pas d'un tel luxe, et lorsqu'on les emploie à des fins éloignées de leur construction initiale on s'expose plus facilement à de mauvaises performances ou autres comportements non-physiques; c'est notamment le cas pour la prédiction des nano-clusters les plus stables, pour les courbes énergie vs. volume ou encore les densités d'états vibrationnels des monocristaux.

Notre approche et celles purement numériques ne s'appliquent évidemment pas aux mêmes échelles et ne cherchent pas à décrire les mêmes propriétés - notamment avec la prise en compte ou non des propriétés électroniques. C'est avant tout une politique conceptuelle très différente qui les motive en amont de toute construction, alors même que les outils manipulés (fonctions de symétrie centrées sur les atomes et réseaux de neurones de haute dimension) peuvent être très similaires. Dans le premier cas, on ne laisse aux outils d'apprentissage automatique qu'un rôle secondaire, à la justification théorique forte et en l'entourant de sécurités afin de prévenir la majorité des aberrations. On garde dans le même temps un accès complet aux états quantiques du système, c'està-dire leur énergie et leur localisation. Même si elles ne sont pas présentées car ce n'est pas le point focal des présents travaux, notre approche nous permet de calculer un grand nombre d'observables inaccessibles aux approches purement paramétriques, par exemple : moment dipolaire, propriétés magnétiques, propriétés optiques. Certaines limitations se font néanmoins remarquer : en particulier, on espérait assurer une bonne représentation des interactions aux plus longues portées en profitant des termes électroniques de DFTB. Ceci nous aurait permis de rendre compte des systèmes non-périodiques, au sein desquels les interactions sont délocalisées; au vu des résultats médiocres obtenus pour les nanoclusters, cet espoir n'est manifestement pas satisfait.

Dans le cas d'outils purement paramétriques, on laisse libre cours à leur flexibilité; ce choix est évidemment à double tranchant, avec les résultats très impressionnants qui s'en accompagnent potentiellement, au prix de l'abandon de la quasi-intégralité de la physique hors du contrôle de l'utilisateur.

Concluons enfin par quelques perspectives : ces travaux se concentrent évidemment sur des directions spécifiques, en laissant de nombreuses à approfondir ou défricher. La performance du potentiel répulsif numérique sur les systèmes isolés est l'une des principales caractéristiques à parfaire, bien que cela semble requérir des modifications conséquentes à notre approche : DFTB est très utilisé dans la communauté pour s'intéresser à des molécules, mais nous n'avons pas favorisé cette direction peu pertinente dans le cas du Silicium
pur. En particulier, le recours à une nouvelle description des environnements atomiques plus adaptée aux systèmes isolés semble inéluctable, bien que l'alliage périodique/nonpériodique reste toujours ardu à couvrir efficacement. En s'intéressant à des systèmes mixtes - typiquement Si-O - on pourrait par ailleurs espérer enrichir fortement la variété de configurations accessibles et éprouver plus encore nos outils. La comparaison avec une quantité accrue de systèmes pertinents physiquement en constitue une motivation supplémentaire. D'un point de vue conceptuel, on peut s'interroger quant à la transférabilité de notre représentation des interactions Si-Si dans des systèmes homonucléaires à ces mêmes interactions dans un système mixte. Dans une temporalité plus courte, l'exploration de nouvelles phases cristallines du Silicium pur semble intéressante : en incluant des cristaux dans des conditions plus extrêmes de pression ou de température, il semble possible d'ébaucher un diagramme de phase global pour cet élément. Le coût des calculs premier principe semble dissuader cette approche : à notre connaissance, aucun digramme de phase complet (incluant à la fois de nombreux polymorphes [118] et des liquides [119]) généré avec DFT ou une autre méthode premier principe n'est disponible pour le Silicium pur, et on se repose usuellement sur la compilation de données expérimentales.

Bibliographie

- E. Schrödinger. An undulatory theory of the mechanics of atoms and molecules. Phys. Rev., 28 :1049–1070, Dec 1926.
- [2] Louis De Broglie. <u>Recherches sur la théorie des quanta</u>. PhD thesis, Migrationuniversité en cours d'affectation, 1924.
- [3] Max Born and Robert Oppenheimer. Zur quantentheorie der molekeln. <u>Annalen</u> der physik, 389(20) :457–484, 1927.
- [4] Robert N Barnett and Uzi Landman. Born-oppenheimer molecular-dynamics simulations of finite systems : Structure and dynamics of (h 2 o) 2. <u>Physical review</u> B, 48(4) :2081, 1993.
- [5] Douglas R Hartree. The wave mechanics of an atom with a non-coulomb central field. part i. theory and methods. In <u>Mathematical Proceedings of the Cambridge</u> Philosophical Society, volume 24, pages 89–110. Cambridge university press, 1928.
- [6] Wolfgang Pauli. The connection between spin and statistics. <u>Physical Review</u>, 58(8):716, 1940.
- [7] J. C. Slater. The theory of complex spectra. Phys. Rev., 34 :1293–1322, Nov 1929.
- [8] Tjalling Koopmans. Über die zuordnung von wellenfunktionen und eigenwerten zu den einzelnen elektronen eines atoms. physica, 1(1-6) :104–113, 1934.
- [9] Christoph Köppl and Hans-Joachim Werner. Parallel and low-order scaling implementation of hartree–fock exchange using local density fitting. Journal of Chemical Theory and Computation, 12(7):3122–3134, 2016. PMID : 27267488.
- [10] Douglas L Strout and Gustavo E Scuseria. A quantitative study of the scaling properties of the hartree–fock method. <u>The Journal of chemical physics</u>, 102(21):8448– 8452, 1995.
- [11] Pierre Hohenberg and Walter Kohn. Inhomogeneous electron gas. <u>Physical review</u>, 136(3B) :B864, 1964.
- [12] Llewellyn H Thomas. The calculation of atomic fields. In <u>Mathematical proceedings</u> of the Cambridge philosophical society, volume 23, pages 542–548. Cambridge University Press, 1927.
- [13] Enrico Fermi. Statistical method to determine some properties of atoms. <u>Rend.</u> <u>Accad. Naz. Lincei, 6(602-607)</u> :5, 1927.
- [14] Walter Kohn and Lu Jeu Sham. Self-consistent equations including exchange and correlation effects. Physical review, 140(4A) :A1133, 1965.

- [15] D. M. Ceperley and B. J. Alder. Ground state of the electron gas by a stochastic method. Phys. Rev. Lett., 45 :566–569, Aug 1980.
- [16] Teepanis Chachiyo. Communication : Simple and accurate uniform electron gas correlation energy for the full range of densities. <u>The Journal of Chemical Physics</u>, 145(2) :021101, 2016.
- [17] K Laasonen, F Csajka, and M Parrinello. Water dimer properties in the gradientcorrected density functional theory. <u>Chemical physics letters</u>, 194(3):172–174, 1992.
- [18] Fiona Sim, Alain St. Amant, Imre Papai, and Dennis R Salahub. Gaussian density functional calculations on hydrogen-bonded systems. <u>Journal of the American</u> Chemical Society, 114(11) :4391–4400, 1992.
- [19] John P. Perdew, J. A. Chevary, S. H. Vosko, Koblar A. Jackson, Mark R. Pederson, D. J. Singh, and Carlos Fiolhais. Atoms, molecules, solids, and surfaces : Applications of the generalized gradient approximation for exchange and correlation. <u>Phys.</u> Rev. B, 46 :6671–6687, Sep 1992.
- [20] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. Phys. Rev. Lett., 77 :3865–3868, Oct 1996.
- [21] Axel D. Becke. Density-functional thermochemistry. iii. the role of exact exchange. The Journal of Chemical Physics, 98(7) :5648–5652, 1993.
- [22] Jianmin Tao, John P Perdew, Viktor N Staroverov, and Gustavo E Scuseria. Climbing the density functional ladder : Nonempirical meta-generalized gradient approximation designed for molecules and solids. <u>Physical Review Letters</u>, 91(14) :146401, 2003.
- [23] G Seifert, H Eschrig, and W Bieger. An approximation variant of lcao-x-alpha methods. Z. Phys. Chem (Leipzig), 267 :529–539, 1986.
- [24] D. Porezag, Th. Frauenheim, Th. Köhler, G. Seifert, and R. Kaschner. Construction of tight-binding-like potentials on the basis of density-functional theory : Application to carbon. Phys. Rev. B, 51 :12947–12957, May 1995.
- [25] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, Th. Frauenheim, S. Suhai, and G. Seifert. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. <u>Phys. Rev. B</u>, 58 :7260–7268, Sep 1998.
- [26] B. Hourahine, B. Aradi, V. Blum, F. Bonafé, A. Buccheri, C. Camacho, C. Cevallos, M. Y. Deshaye, T. Dumitrică, A. Dominguez, S. Ehlert, M. Elstner, T. van der Heide, J. Hermann, S. Irle, J. J. Kranz, C. Köhler, T. Kowalczyk, T. Kubař, I. S. Lee, V. Lutsker, R. J. Maurer, S. K. Min, I. Mitchell, C. Negre, T. A. Niehaus, A. M. N. Niklasson, A. J. Page, A. Pecchia, G. Penazzi, M. P. Persson, J. Řezáč, C. G. Sánchez, M. Sternberg, M. Stöhr, F. Stuckenberg, A. Tkatchenko, V. W.-z. Yu, and T. Frauenheim. Dftb+, a software package for efficient approximate density functional theory based atomistic simulations. <u>The Journal of Chemical Physics</u>, 152(12) :124101, 2020.
- [27] Pekka Koskinen and Ville Mäkinen. Density-functional tight-binding for beginners. Computational Materials Science, 47(1):237–253, 2009.

- [28] Marcus Elstner and Gotthard Seifert. Density functional tight binding. Philosophical Transactions of the Royal Society A : Mathematical, Physical and Engineering Sciences, 372(2011) :20120483, 2014.
- [29] Michael Gaus, Qiang Cui, and Marcus Elstner. Dftb3 : Extension of the selfconsistent-charge density-functional tight-binding method (scc-dftb). Journal of <u>Chemical Theory and Computation</u>, 7(4) :931–948, 2011.
- [30] Thomas Frauenheim, Gotthard Seifert, Marcus Elstner, Thomas Niehaus, Christof Köhler, Marc Amkreutz, Michael Sternberg, Zoltán Hajnal, Aldo Di Carlo, and Sándor Suhai. Atomistic simulations of complex materials : ground-state and excitedstate properties. Journal of Physics : Condensed Matter, 14(11) :3015, 2002.
- [31] Michael Sternberg. The atomic structure of diamond surfaces and interfaces. <u>Univ.</u> Paderborn, 2001.
- [32] John C Slater and George F Koster. Simplified leao method for the periodic potential problem. Physical Review, 94(6) :1498, 1954.
- [33] Rudolph Pariser. Theory of the electronic spectra and structure of the polyacenes and of alternant hydrocarbons. <u>The Journal of Chemical Physics</u>, 24(2) :250–268, 1956.
- [34] Marcus Elstner. The scc-dftb method and its application to biological systems. Theoretical Chemistry Accounts, 116(1):316–325, 2006.
- [35] Marcus Elstner. The scc-dftb method and its application to biological systems. Theoretical Chemistry Accounts, 116(1):316–325, 2006.
- [36] Alexander Sieck. Structure and physical properties of silicon clusters and of vacancy clusters in bulk silicon. Univ. Paderborn, 2000.
- [37] Akshay Krishna Ammothum Kandy, Eddie Wadbro, Bálint Aradi, Peter Broqvist, and Jolla Kullgren. Curvature constrained splines for dftb repulsive potential parametrization. <u>Journal of Chemical Theory and Computation</u>, 17(3) :1771–1781, 2021.
- [38] Maxime Van den Bossche, Henrik Grönbeck, and Bjørk Hammer. Tight-binding approximation-enhanced global optimization. <u>Journal of chemical theory and</u> computation, 14(5) :2797–2807, 2018.
- [39] Haichen Li, Christopher Collins, Matteus Tanha, Geoffrey J Gordon, and David J Yaron. A density functional tight binding layer for deep learning of chemical hamiltonians. Journal of chemical theory and computation, 14(11):5764–5776, 2018.
- [40] Nir Goldman, Bálint Aradi, Rebecca K Lindsey, and Laurence E Fried. Development of a multicenter density functional tight binding model for plutonium surface hydriding. <u>Journal of chemical theory and computation</u>, 14(5) :2652–2660, 2018.
- [41] P.L. Tchebychev. <u>Théorie des mécanismes connus sous le nom de parallélogrammes</u>. Imprimerie de l'Académie impériale des sciences, 1853.
- [42] Nir Goldman, Kyoung E Kweon, Babak Sadigh, Tae Wook Heo, Rebecca K Lindsey, C Huy Pham, Laurence E Fried, Bálint Aradi, Kiel Holliday, Jason R Jeffries, et al. Semi-automated creation of density functional tight binding models through leveraging chebyshev polynomial-based force fields. Journal of Chemical Theory and Computation, 2021.

- [43] Simon Wengert, Gábor Csányi, Karsten Reuter, and Johannes T. Margraf. Dataefficient machine learning for molecular crystal structure prediction. <u>Chem. Sci.</u>, 12 :4536–4546, 2021.
- [44] Julian J Kranz, Maximilian Kubillus, Raghunathan Ramakrishnan, O Anatole von Lilienfeld, and Marcus Elstner. Generalized density-functional tight-binding repulsive potentials from unsupervised machine learning. Journal of chemical theory and computation, 14(5) :2341–2352, 2018.
- [45] Chiara Panosetti, Artur Engelmann, Lydia Nemec, Karsten Reuter, and Johannes T. Margraf. Learning to use the force : Fitting repulsive potentials in density-functional tight-binding with gaussian process regression. Journal of Chemical Theory and Computation, 16(4) :2181–2191, 2020. PMID : 32155065.
- [46] Martin Stöhr, Leonardo Medrano Sandonas, and Alexandre Tkatchenko. Accurate many-body repulsive potentials for density-functional tight binding from deep tensor neural networks. <u>The Journal of Physical Chemistry Letters</u>, 11(16):6835–6843, 2020. PMID : 32787209.
- [47] Volker L Deringer, Miguel A Caro, and Gábor Csányi. Machine learning interatomic potentials as emerging tools for materials science. <u>Advanced Materials</u>, 31(46) :1902765, 2019.
- [48] Jörg Behler. Four generations of high-dimensional neural network potentials. Chemical Reviews, 2021.
- [49] Tsz Wai Ko, Jonas A Finkler, Stefan Goedecker, and Jörg Behler. Generalpurpose machine learning potentials capturing nonlocal charge transfer. <u>Accounts</u> of Chemical Research, 54(4) :808–817, 2021.
- [50] Thomas Bayes. Lii. an essay towards solving a problem in the doctrine of chances. by the late rev. mr. bayes, frs communicated by mr. price, in a letter to john canton, amfr s. <u>Philosophical transactions of the Royal Society of London</u>, pages 370–418, 1763.
- [51] Pierre-Simon de Laplace. <u>Théorie analytique des probabilités; Par M. le comte Laplace,...</u> Mme Ve Courcier, imprimeur-libraire pour les mathématiques et la marine ..., 1812.
- [52] Alan Mathison Turing. Intelligent machinery, 1948.
- [53] Marvin Lee Minsky. <u>Theory of neural-analog reinforcement systems and its</u> application to the brain-model problem. PhD thesis, Princeton University, 1954.
- [54] Mark R Baker and Rajendra B Patil. Universal approximation theorem for interval neural networks. Reliable Computing, 4(3) :235–239, 1998.
- [55] Frank Rosenblatt. Perceptron simulation experiments. <u>Proceedings of the IRE</u>, 48(3):301–309, 1960.
- [56] Kunihiko Fukushima and Sei Miyake. Neocognitron : A self-organizing neural network model for a mechanism of visual pattern recognition. In <u>Competition and</u> cooperation in neural nets, pages 267–285. Springer, 1982.
- [57] Raul Rojas. The backpropagation algorithm. In <u>Neural networks</u>, pages 149–182. Springer, 1996.

- [58] DA Hopfield. Diagonal recurrent neural networks for dynamics control. J Proc Nat Acad Sci USA, 79 :2554–2561, 1982.
- [59] Christopher John Cornish Hellaby Watkins. Learning from delayed rewards. <u>thesis</u>, 1989.
- [60] Tin Kam Ho. Random decision forests. In <u>Proceedings of 3rd International</u> <u>Conference on Document Analysis and Recognition</u>, volume 1, pages 278–282 vol.1, 1995.
- [61] Corinna Cortes and Vladimir Vapnik. Support-vector networks. <u>Machine learning</u>, 20(3) :273–297, 1995.
- [62] Thomas Hofmann, Bernhard Schölkopf, and Alexander J Smola. Kernel methods in machine learning. The annals of statistics, pages 1171–1220, 2008.
- [63] Arthur L Samuel. Some studies in machine learning using the game of checkers. IBM Journal of research and development, 3(3) :210–229, 1959.
- [64] Donald Michie. Experiments on the Mechanization of Game-Learning Part I. Characterization of the Model and its parameters. <u>The Computer Journal</u>, 6(3) :232– 236, 11 1963.
- [65] Murray Campbell, A Joseph Hoane Jr, and Feng-hsiung Hsu. Deep blue. <u>Artificial</u> intelligence, 134(1-2) :57–83, 2002.
- [66] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, et al. Mastering the game of go without human knowledge. <u>nature</u>, 550(7676):354– 359, 2017.
- [67] Terrence J Sejnowski and Charles R Rosenberg. Parallel networks that learn to pronounce english text. <u>Complex systems</u>, 1(1) :145–168, 1987.
- [68] Quoc V Le. Building high-level features using large scale unsupervised learning. In <u>2013 IEEE international conference on acoustics</u>, speech and signal processing, pages 8595–8598. IEEE, 2013.
- [69] Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf. Deepface : Closing the gap to human-level performance in face verification. In <u>Proceedings of</u> the IEEE conference on computer vision and pattern recognition, pages 1701–1708, 2014.
- [70] John E Lennard-Jones. Cohesion. <u>Proceedings of the Physical Society (1926-1948)</u>, 43(5):461, 1931.
- [71] Albert P. Bartók, Mike C. Payne, Risi Kondor, and Gábor Csányi. Gaussian approximation potentials : The accuracy of quantum mechanics, without the electrons. Phys. Rev. Lett., 104 :136403, Apr 2010.
- [72] Albert P Bartók, Michael J Gillan, Frederick R Manby, and Gábor Csányi. Machinelearning approach for one-and two-body corrections to density functional theory : Applications to molecular and condensed water. <u>Physical Review B</u>, 88(5) :054104, 2013.
- [73] John F Stanton. Why ccsd (t) works : a different perspective. <u>Chemical Physics</u> Letters, 281(1-3) :130–134, 1997.

- [74] Albert P Bartók, James Kermode, Noam Bernstein, and Gábor Csányi. Machine learning a general-purpose interatomic potential for silicon. <u>Physical Review X</u>, 8(4):041048, 2018.
- [75] Albert P Bartók and Gábor Csányi. Gaussian approximation potentials : A brief tutorial introduction. <u>International Journal of Quantum Chemistry</u>, 115(16) :1051– 1057, 2015.
- [76] Xiao-Xiao Niu and Ching Y Suen. A novel hybrid cnn-svm classifier for recognizing handwritten digits. Pattern Recognition, 45(4) :1318–1325, 2012.
- [77] Jörg Behler and Michele Parrinello. Generalized neural-network representation of high-dimensional potential-energy surfaces. <u>Physical review letters</u>, 98(14) :146401, 2007.
- [78] Jörg Behler. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. The Journal of chemical physics, 134(7):074106, 2011.
- [79] J. Tersoff. Modeling solid-state chemistry : Interatomic potentials for multicomponent systems. Phys. Rev. B, 39 :5566–5568, Mar 1989.
- [80] Albert P Bartók, Risi Kondor, and Gábor Csányi. On representing chemical environments. Physical Review B, 87(18) :184115, 2013.
- [81] Ya A Smorodinskiĭ and Leonid A Shelepin. Clebsch-gordan coefficients, viewed from different sides. Soviet Physics Uspekhi, 15(1) :1, 1972.
- [82] Risi Kondor. A complete set of rotationally and translationally invariant features for images. CoRR, abs/cs/0701127, 2007.
- [83] Matthias Rupp, Alexandre Tkatchenko, Klaus-Robert Müller, and O. Anatole von Lilienfeld. Fast and accurate modeling of molecular atomization energies with machine learning. Phys. Rev. Lett., 108 :058301, Jan 2012.
- [84] Jörg Behler. Constructing high-dimensional neural network potentials : A tutorial review. International Journal of Quantum Chemistry, 115(16) :1032–1050, 2015.
- [85] Rudolph Emil Kalman et al. A new approach to linear filtering and prediction problems [j]. Journal of basic Engineering, 82(1):35–45, 1960.
- [86] Andreas Singraber, Tobias Morawietz, Jörg Behler, and Christoph Dellago. Parallel multistream training of high-dimensional neural network potentials. Journal of chemical theory and computation, 15(5) :3075–3092, 2019.
- [87] Mohammad Wahiduzzaman, Augusto F Oliveira, Pier Philipsen, Lyuben Zhechkov, Erik Van Lenthe, Henryk A Witek, and Thomas Heine. Dftb parameters for the periodic table : Part 1, electronic structure. <u>Journal of chemical theory and</u> computation, 9(9) :4006–4017, 2013.
- [88] Jan Gerit Brandenburg, Manuel Hochheim, Thomas Bredow, and Stefan Grimme. Low-cost quantum chemical methods for noncovalent interactions. <u>The journal of</u> physical chemistry letters, 5(24) :4275–4284, 2014.
- [89] Arnaud Fihey and Denis Jacquemin. Performances of density functional tightbinding methods for describing ground and excited state geometries of organic molecules. Journal of chemical theory and computation, 15(11) :6267–6276, 2019.

- [90] Shima Heydariyan, Mohammad Reza Nouri, Mojtaba Alaei, Zahed Allahyari, and Thomas A Niehaus. New candidates for the global minimum of medium-sized silicon clusters : A hybrid dftb/dft genetic algorithm applied to si n, n= 8-80. <u>The Journal</u> of chemical physics, 149(7) :074313, 2018.
- [91] Georg Kresse and Jürgen Hafner. Ab initio molecular dynamics for liquid metals. Physical Review B, 47(1) :558, 1993.
- [92] Georg Kresse and Jürgen Furthmüller. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. <u>Computational</u> materials science, 6(1) :15–50, 1996.
- [93] Georg Kresse and Jürgen Furthmüller. Efficient iterative schemes for ab initio totalenergy calculations using a plane-wave basis set. <u>Physical review B</u>, 54(16) :11169, 1996.
- [94] P. E. Blöchl. Projector augmented-wave method. <u>Phys. Rev. B</u>, 50 :17953–17979, Dec 1994.
- [95] Georg Kresse and Daniel Joubert. From ultrasoft pseudopotentials to the projector augmented-wave method. Physical review b, 59(3) :1758, 1999.
- [96] L. Boltzmann. <u>Studien über das Gleichgewicht der lebendigen Kraft zwischen</u> bewegten materiellen Punkten. k. und k. Hof- und Staatsdr., 1868.
- [97] DS Wilks. Indices of rank histogram flatness and their sampling properties. <u>Monthly</u> Weather Review, 147(2) :763–769, 2019.
- [98] Yunxing Zuo, Chi Chen, Xiangguo Li, Zhi Deng, Yiming Chen, Jörg Behler, Gábor Csányi, Alexander V Shapeev, Aidan P Thompson, Mitchell A Wood, et al. Performance and cost assessment of machine learning interatomic potentials. <u>The</u> Journal of Physical Chemistry A, 124(4) :731–745, 2020.
- [99] Stewart J Clark, Matthew D Segall, Chris J Pickard, Phil J Hasnip, Matt IJ Probert, Keith Refson, and Mike C Payne. First principles methods using castep. Zeitschrift für Kristallographie-Crystalline Materials, 220(5-6):567–570, 2005.
- [100] John P. Perdew and Yue Wang. Accurate and simple analytic representation of the electron-gas correlation energy. Phys. Rev. B, 45 :13244–13249, Jun 1992.
- [101] Akshay Krishna Ammothum Kandy, Eddie Wadbro, Bálint Aradi, Peter Broqvist, and Jolla Kullgren. Curvature constrained splines for dftb repulsive potential parametrization. Journal of Chemical Theory and Computation, 17(3) :1771–1781, 2021. PMID : 33606527.
- [102] A Togo and I Tanaka. First principles phonon calculations in materials science. Scr. Mater., 108 :1–5, Nov 2015.
- [103] Hendrik J Monkhorst and James D Pack. Special points for brillouin-zone integrations. Physical review B, 13(12) :5188, 1976.
- [104] Koji Sueoka, Eiji Kamiyama, and Jan Vanhellemont. Density functional theory study on the impact of heavy doping on si intrinsic point defect properties and implications for single crystal growth from a melt. Journal of Applied Physics, 114(15) :153510, 2013.

- [105] Fabiano Corsetti and Arash A Mostofi. System-size convergence of point defect properties : The case of the silicon vacancy. <u>Physical Review B</u>, 84(3) :035209, 2011.
- [106] L Ferry, F Virot, M Barrachin, Y Ferro, C Pardanaud, D Matveev, M Wensing, T Dittmar, M Koppen, and C Linsmeier. Theoretical investigation on the point defect formation energies in beryllium and comparison with experiments. <u>Nuclear</u> Materials and Energy, 12 :453–457, 2017.
- [107] TS Bush, C Rchard A Catlow, and PD Battle. Evolutionary programming techniques for predicting inorganic crystal structures. <u>Journal of Materials Chemistry</u>, 5(8) :1269–1272, 1995.
- [108] Artem R Oganov, Andriy O Lyakhov, and Mario Valle. How evolutionary crystal structure prediction works and why. <u>Accounts of chemical research</u>, 44(3):227–237, 2011.
- [109] Volker Blum, Ralf Gehrke, Felix Hanke, Paula Havu, Ville Havu, Xinguo Ren, Karsten Reuter, and Matthias Scheffler. Ab initio molecular simulations with numeric atom-centered orbitals. <u>Computer Physics Communications</u>, 180(11) :2175–2196, 2009.
- [110] Magnus Rudolph Hestenes, Eduard Stiefel, et al. <u>Methods of conjugate gradients</u> for solving linear systems, volume 49. NBS Washington, DC, 1952.
- [111] Martin F. Jarrold and Vladimir A. Constant. Silicon cluster ions : Evidence for a structural transition. Phys. Rev. Lett., 67 :2994–2997, Nov 1991.
- [112] Y Waseda and K Suzuki. Structure of molten silicon and germanium by x-ray diffraction. Zeitschrift für Physik B Condensed Matter, 20(4) :339–343, 1975.
- [113] Khalid Laaziri, S Kycia, S Roorda, M Chicoine, JL Robertson, J Wang, and SC Moss. High resolution radial distribution function of pure amorphous silicon. Physical review letters, 82(17) :3460, 1999.
- [114] MMJ Treacy and KB Borisenko. The local structure of amorphous silicon. <u>Science</u>, 335(6071) :950–953, 2012.
- [115] J Fortner and JS Lannin. Radial distribution functions of amorphous silicon. Physical Review B, 39(8) :5527, 1989.
- [116] Jerry Tersoff. Empirical interatomic potential for silicon with improved elastic properties. Physical Review B, 38(14) :9902, 1988.
- [117] W. A. Kamitakahara, H. R. Shanks, J. F. McClelland, U. Buchenau, F. Gompf, and L. Pintschovius. Measurement of phonon densities of states for pure and hydrogenated amorphous silicon. <u>Phys. Rev. Lett.</u>, 52 :644–647, Feb 1984.
- [118] Cong Li, Cuiping Wang, Jiajia Han, Lihui Yan, Bin Deng, and Xingjun Liu. A comprehensive study of the high-pressure-temperature phase diagram of silicon. Journal of materials science, 53(10) :7475–7485, 2018.
- [119] M. Kaczmarski, O. N. Bedoya-Martínez, and E. R. Hernández. Phase diagram of silicon from atomistic simulations. <u>Phys. Rev. Lett.</u>, 94 :095701, Mar 2005.
- [120] Giulio Imbalzano, Andrea Anelli, Daniele Giofré, Sinja Klees, Jörg Behler, and Michele Ceriotti. Automatic selection of atomic fingerprints and reference configurations for machine-learning potentials. <u>The Journal of chemical physics</u>, 148(24) :241730, 2018.

Annexe A

Choix des descripteurs

Nous nous intéressons ici aux approches possibles pour sélectionner des fonctions de symétrie. Évidemment, un jeu de fonctions de symétrie n'est que rarement *bon* ou *mauvais* en soi ; il faut s'intéresser à sa capacité de représentation pour un jeu de données de référence (c'est-à-dire d'environnements atomiques) donné.

A.1 Génération de jeux de paramètres

Avant de pouvoir sélectionner d'une manière ou d'une autre des fonctions de symétrie, il faut tout d'abord avoir un jeu de base au sein duquel opérer la sélection. Deux approches sont envisageables, selon le degré d'automatisation que l'on souhaite adopter ; cependant, le choix de l'utilisateur est *a minima* requis pour déterminer le *rayon de coupure maximal* $r_{\text{cut}}^{\text{max}}$.

Ce rayon de coupure maximal détermine la distance au-delà de laquelle on ignore toute interaction. Il est évidemment très dépendant de l'espèce chimique considérée; il l'est tout autant de l'application visée.

Si l'on cherche à construire un potentiel numérique (voir 5.2), on peut s'inspirer des potentiels empiriques déjà utilisés dans la littérature pour des applications et systèmes similaires : la plupart d'entre eux ont un rayon de coupure.

Il est à noter qu'augmenter le rayon de coupure améliore souvent la capacité de représentation d'une méthode. Si l'importance des contributions est en général plus faible aux longues distances, elles ne sont pas nulles : plus le rayon de coupure est élevé, plus on tient compte d'informations sur le système, et le risque de négliger une contribution importante s'amoindrit. Cependant, augmenter le rayon de coupure impacte fortement le temps de calcul; on tient compte de beaucoup plus de voisins, ce qui implique beaucoup plus de calculs de distances, de moins en moins influents à mesure que l'on s'éloigne de l'atome central. L'introduction du rayon de coupure sert, historiquement, à pallier ce problème en trouvant un compromis entre efficacité de calcul et prise en compte des contributions importantes. Dans le cadre de l'application qui nous intéresse, c'est-à-dire la génération de potentiels répulsifs $E_{\rm rep}$ pour la méthode DFTB, la justification d'un rayon de coupure est toute autre. La distance caractéristique est typiquement beaucoup plus courte, car les interactions répulsives s'étendent théoriquement à peine plus loin que les premiers voisins. Un potentiel répulsif est dans les faits construit comme un terme *«correctif»* : il aspire à rendre compte de la différence entre l'énergie DFT et l'énergie DFTB sans potentiel répulsif. Si cette différence est en principe uniquement constituée d'interactions répulsives de courte portée, les approximations de DFTB impliquent quelques imprécisions aux plus longues distances. En conséquence, plus on augmente le rayon de coupure, plus il est probable que l'on représente autre chose que des interactions purement répulsives. Bien que cela permette d'imiter plus précisément DFT, on réduit par la même occasion la transférabilité du potentiel (voir 3.3).

Ici encore, il est possible de s'appuyer sur les représentations existantes des interactions répulsives : les potentiels de paire utilisées usuellement ont déjà un rayon de coupure, typiquement $r_{\rm cut} \in [2;3]$ Å. En connaissant la gamme de système de référence que l'on souhaite représenter, l'idéal est néanmoins de choisir $r_{\rm cut}$ de manière à inclure tous les premiers voisins.

Une fois le rayon de coupure choisi, on peut générer des jeux de paramètres de fonctions de symétrie. D'instinct, plusieurs possibilités s'offrent à nous : utiliser une intuition physique, automatiser la génération de données ou allier ces deux approches.

Assez peu de détails généraux peuvent être donnés concernant la première approche. Si l'on connaît le système à étudier, on peut jouer avec le paramètre r_s des fonctions G_2 pour assurer une sensibilité maximale aux distances clés. De même, jouer avec la résolution angulaire ζ ou le paramètre λ des fonctions G_4 peut permettre une sensibilité maximale autour d'angles d'intérêt.

À l'inverse, une approche systématique pour générer des fonctions de symétrie quadrillant régulièrement les distances et angles de l'environnement est proposée par G. Imbalzano *et. al*, [120].

Pour les fonctions G_2 , on peut distinguer deux catégories. La première est constituée de fonctions de symétrie radiales centrées sur l'atome central de l'environnement, c'està-dire pour lesquelles $r_s = 0$ Å. Il reste alors à choisir des valeurs de η pertinentes. Une petite valeur de η implique une fonction G_2 large, s'étendant jusqu'aux grandes distances mais aux variations peu sensibles aux petits changements de distance. À l'inverse, de grandes valeurs de η sont synonymes de fonctions G_2 fines, très sensibles aux variations aux courtes distances. Pour l'application qui nous intéresse, ce genre de comportement est capital pour assurer une sensibilité forte aux courtes distances afin de rendre détectable tout rapprochement excessif d'atomes.

On génère donc n + 1 fonctions de symétrie G_2 où $r_s = 0$ comme :

$$\eta_m = \left(\frac{n^{m/n}}{r_{\rm cut}}\right)^2 \tag{A.1}$$

Concernant les fonctions G_2 où $r_s \neq 0$ Å, l'idée est de centrer les gaussiennes à des

intervalles de distances réguliers.

$$r_{s,m} = \frac{r_{\rm cut}}{n^{m/n}}.$$

On choisit la largeur des gaussiennes de manière à avoir des fonctions fines aux courtes distances et qui s'élargissent à mesure que l'on s'approche du rayon de coupure.

$$\eta_m = \frac{1}{\left(r_{s,n-m} - r_{s,n-m-1}\right)^2}.$$

On donne un exemple d'utilisation de cette approche dans la figure A.1, où l'on génère 17 fonctions G_2 (n = 8) pour $r_{\text{cut}} = 5.0$ Å.



FIGURE A.1 – Fonctions de symétrie radiales générées automatiquement.

Pour les fonctions G_4 , il faut obtenir des jeux de paramètres $\{\eta, \lambda, \zeta\}$, r_{cut} étant fixé par l'utilisateur.

- On génère η de la même manière que pour les fonctions G_2 centrées en 0 (A.1).
- λ ne peut prendre que deux valeurs : -1 et +1. Pour chaque autre paramètre fixé, on a donc deux fonctions G_4 avec $\lambda = \pm 1$.
- On choisit plusieurs valeurs de ζ sur une échelle logarithmique afin de quadriller convenablement la distribution angulaire du système (voir Fig. 5.11), typiquement ζ = {1, 4, 16, · · · }.

A.2 Sélection préalable

Une fois que l'on dispose de plusieurs fonctions de symétrie, il peut s'avérer capital de sélectionner celles qui sont les plus adaptées à notre application.

$$\left\{\vec{G}_i\right\}_{i=1}^{N_{\text{init}}^{\text{ACSF}}} \to \left\{\vec{G}_i\right\}_{i=1}^{N_{\text{sel}}^{\text{ACSF}}}.$$

Une première sélection est possible en amont de tout processus d'apprentissage, en identifiant pour un jeu d'environnements atomiques de référence donné les fonctions de symétrie les plus pertinentes.

La pertinence d'une fonction de symétrie par rapport aux données de référence peut être évaluée individuellement : quelle plage de valeurs la fonction adopte-t-elle? Si une fonction reste quasiment constante pour l'intégralité du jeu de données de référence, son intérêt est moindre car elle ne permet pas de discriminer facilement entre deux environnements. En conséquence, il est intéressant sélectionner des fonctions de symétrie si l'écart entre leurs valeurs maximales et minimales est supérieur à un certain seuil.

$$P: \left\{\vec{G}_{i}\right\}_{i=1}^{N_{\text{init}}^{\text{ACSF}}} \mapsto \left\{\vec{G}_{i}\right\}_{i=1}^{N_{\text{sel}}^{\text{ACSF}}} P\left(\left\{\vec{G}_{i}\right\}_{i=1}^{N_{\text{init}}^{\text{ACSF}}}\right) = \left\{\vec{G}_{i}\right\}, |\max(G_{i}) - \min(G_{i})| > \sigma.$$
(A.2)

Il n'existe pas de valeur optimale pour ce seuil : on s'en sert surtout pour éliminer les fonctions de symétrie retournant des valeurs nulles su l'intégralité des données de référence. Dans la pratique, une valeur de l'ordre $\sigma \geq 1 \times 10^{-6}$ remplit correctement ce rôle, mais on peut si besoin augmenter ce seuil de manière à supprimer en amont de tout apprentissage une quantité plus grande de fonctions de symétrie.

On peut également évaluer la pertinence d'une fonction de symétrie appliquée au jeu de données de référence par rapport aux fonctions de symétrie déjà sélectionnées : si deux fonctions de symétrie appliquées au jeu de données de référence sont très similaires, on peut se contenter de n'en sélectionner qu'une des deux.

Pour un jeu d'environnements atomiques {**R**} de référence, l'application d'une fonction de symétrie G_i nous permet d'obtenir les éléments de vecteur descripteur associés $\{\vec{D}_i = \vec{G}_i(\mathbf{R})\}.$

On peut à partir de là calculer une distance entre plusieurs vecteurs \vec{D} . Une première approche consiste donc à sélectionner, itérativement, la fonction de symétrie G_i telle que la distance avec les autres fonctions sélectionnées est maximale (A.3) : on nomme cette méthode *Farthest Point Sampling* (FPS - échantillonnage au point le plus distant).

$$i = \operatorname{argmax}\left(\min_{j} \left| \vec{D}_{i} - \vec{D}_{j} \right| \right).$$
(A.3)

Ici, j désigne toutes les fonctions de symétrie déjà sélectionnées.

Enfin, la méthode de la corrélation de *Pearsons* propose, tout comme la méthode FPS, de sélectionner les fonctions de symétrie d'après leurs différences relatives. Ici, au lieu de calculer une distance entre les fonctions de symétrie G_i , on calcule leurs corrélations respectives. On définit cette corrélation P_{ij} pour deux fonctions de symétrie G_i et G_j comme :

$$P_{ij} = \frac{\sum_{k=1}^{N^{env}} \left(G_i \left(\mathbf{R}_k \right) - \bar{G}_i \right) \cdot \left(G_j \left(\mathbf{R}_k \right) - \bar{G}_j \right)}{N^{env} \cdot \sigma \left(G_i \right) \cdot \sigma \left(G_j \right)},$$
(A.4)

où

- $G_i(\mathbf{R}_k)$ est la valeur calculée par la $i^{\text{ème}}$ fonction de symétrie pour le $k^{\text{ème}}$ envi-
- ronnement. $\bar{G}_i = \frac{1}{N^{\text{env}}} \sum_{k=1}^{N^{\text{env}}} G_i(\mathbf{R}_k)$ est la valeur moyenne calculée par la $i^{\text{ème}}$ fonction de
- $\sigma(G_i)$ est l'écart-type des valeurs calculées par la $i^{\text{ème}}$ fonction de symétrie.

On peut imaginer plusieurs autres méthodes de sélection de descripteurs [120], comme par exemple la méthode de la décomposition CUR; l'idée ici n'est pas d'être exhaustif mais de présenter la philosophie générale derrière ces méthodes.

Sélection après apprentissage A.3

La sélection des fonctions de symétrie les plus pertinentes peut également se faire après l'optimisation d'une méthode d'apprentissage automatique - dans notre cas, un réseau de neurones.

Cette démarche s'inscrit dans une logique plus proche de l'optimisation d'hyperparamètres, fréquente lorsque l'on construit un réseau de neurones.

On définit un hyperparamètre comme un paramètre qui est fixé par l'utilisateur pour contrôler la méthode : celui-ci n'est pas optimisé par le processus d'apprentissage (à l'inverse des *poids* et *biais*). On peut par exemple citer la structure du réseau de neurones (nombre de couches cachées, nombre de neurones par couches cachées), mais aussi la fonction de transfert utilisée ou l'algorithme d'apprentissage. Tous ces facteurs influent sur le réseau, de ses performances numériques à sa capacité de prédiction.

Usuellement, ces hyperparamètres sont ajustés par processus d'essai-erreur. Initialement, on prépare un apprentissage en leur attribuant des valeurs données, que l'on éprouve en optimisant le réseau de neurones. Si les propriétés de l'outil optimisé ne sont pas satisfaisantes, on fait varier un ou plusieurs hyperparamètres (ex. : on ajoute une couche cachée), on teste à nouveau en ré-optimisant le nouveau réseau de neurones et on évalue l'effet des modifications apportées (ex. : plus de flexibilité mais apparition de sur-apprentissage). On répète ces opérations jusqu'à atteindre un résultat satisfaisant.

Afin de sélectionner des fonctions de symétrie, on peut effectuer le même genre de

manipulations. Dans un premier temps, on en choisit un nombre conséquent et on optimise un réseau de neurones; le processus est long, et le réseau obtenu n'est pas utilisable tel quel en raison du nombre absurdement élevé de calculs nécessaires pour évaluer toutes les fonctions de symétrie.

On peut en revanche identifier les fonctions dont la valeur importe peu au réseau optimisé, c'est-à-dire celles dont les variations modifient le vecteur de sortie \vec{Y} du réseau moins qu'une valeur seuil.

$$\frac{\frac{\partial Y}{\partial G_i}}{\sum_{j=1}^{N_{\text{ASCF}}} \frac{\partial \vec{Y}}{\partial G_j}} < \sigma,$$

Remarque : la valeur seuil à employer est évidemment dépendante de N_{ACSF} , le nombre de fonction de symétrie. Par exemple, si une fonction modifie d'au 1% la valeur de sortie d'un réseau employant 10 fonctions de symétrie, elle est *a priori* moins utile au résultat final que dans le cas où une fonction modifie de 1% la valeur de sortie d'un réseau en utilisant plusieurs centaines.

Après suppression des fonctions de symétrie *«inutiles»*, on peut réitérer le processus jusqu'à en avoir retenu un nombre satisfaisant : suffisamment pour représenter le jeu de données de référence de manière acceptable, et suffisamment peu pour éviter toute redondance et permettre un temps de calcul efficace.

Annexe B

Calcul des forces répulsives

En simulation atomistique, il est capital d'avoir accès aux forces exercées sur les atomes pour de nombreuses applications. Par exemple, l'optimisation de la géométrie d'un système donné cherche à déterminer un agencement des atomes du système tel que les forces agissant sur ces derniers soient suffisamment petites. De même, un calcul de dynamique moléculaire vise à simuler le mouvement physique des atomes pour analyser leur comportement et extraire des propriétés physiques dynamiques. Pour cela, il est évidemment nécessaire de connaître les forces agissant sur les atomes qui déterminent leurs trajectoires.

Par définition, la force \vec{F}_i agissant sur un atome *i* représente la variation de l'énergie totale du système induite par un mouvement de cet atome (B.1). Cette force est donc un vecteur orienté dans l'espace; on peut lui définir des composantes dans les trois dimensions spatiales $[F_{i,x}, F_{i,y}, F_{i,z}]$, liées chacune à la variation d'énergie totale induite par le déplacement de l'atome *i* dans les trois directions de l'espace $\vec{x}, \vec{y}, \vec{z}$.

$$F_{i,\alpha} = -\frac{\partial E_{\text{tot}}}{\partial r_{i,\alpha}} , \ \alpha = \{x, y, z\}.$$
(B.1)

Pour rappel (3.18), on peut exprimer les forces DFTB $F_{i,\alpha}^{\text{DFTB}}$ comme :

$$F_{i,\alpha} = -\sum_{e=1}^{N_e} \sum_j \sum_{\mu \in i} \sum_{\nu \in j} f_e c_{e,\mu} c_{e\nu} \left[\frac{\partial H^0_{\mu\nu}}{\partial r_{i,\alpha}} - \left(\varepsilon_e - h^1_{\mu\nu} \right) \frac{\partial S_{\mu\nu}}{\partial r_{i,\alpha}} \right] - \Delta q_i \sum_j \frac{\partial \gamma_{ij}}{\partial r_{i,\alpha}} \Delta q_j - \frac{\partial E^{\text{boite}}_{\text{rep}}}{\partial r_{i,\alpha}},$$

que l'on peut réécrire

$$F_{i,\alpha}^{\rm DFTB} = -\frac{\partial E_{\rm DFTB}^{\rm boite} \left[n_0(\vec{r}), \delta n(\vec{r})\right]}{\partial r_{i,\alpha}} = -\frac{\partial E_{\rm rep}^{\rm boite}}{\partial r_{i,\alpha}} - \frac{\partial E_{\rm DFTB, E_{\rm rep}}^{\rm boite}}{\partial r_{i,\alpha}} = F_{i,\alpha}^{\rm rep} + F_{i,\alpha}^{\rm DFTB}, E_{\rm rep}$$

La grandeur $F_{i,\alpha}^{\text{DFTB},E_{\text{Teg}}}$ est déjà calculée par DFTB et est indépendante des interactions répulsives. Déterminer les forces agissant sur les atomes nous demande donc d'évaluer $F_{i,\alpha}^{\text{rep}}$ pour tous les atomes *i* de la boîte de simulation (B.2).

$$F_{i,\alpha}^{\rm rep} = -\frac{\partial E_{\rm rep}^{\rm boite}}{\partial r_{i,\alpha}}.$$
 (B.2)

Dans le cadre de nos travaux, on rappelle que

$$E_{\mathrm{rep}}^{\mathrm{boite,NN}} = \sum_{j=1}^{N_{\mathrm{atomes}}^{\mathrm{boite}}} E_{\mathrm{rep}}^{j,\mathrm{NN}}$$

d'où

$$F_{i,\alpha}^{\text{rep,NN}} = -\sum_{j=1}^{N_{\text{atomes}}^{\text{boite}}} \frac{\partial E_{\text{rep}}^{j,\text{NN}}}{\partial r_{i,\alpha}}.$$
(B.3)

En utilisant (6.1), on peut appliquer le théorème de dérivation des fonctions composées et écrire :

$$F_{i,\alpha}^{\text{rep,NN}} = -\sum_{j=1}^{N_{\text{atomes}}^{\text{boite}}} \frac{\partial \text{NN}\left(G(\mathbf{R}_{j})\right)}{\partial r_{i,\alpha}} = -\sum_{j=1}^{N_{\text{atomes}}^{\text{boite}}} \frac{\partial \text{NN}\left(G(\mathbf{R}_{j})\right)}{\partial G(\mathbf{R}_{j})} \cdot \frac{\partial G(\mathbf{R}_{j})}{\partial r_{i,\alpha}}.$$
 (B.4)

Il semble donc que le calcul des forces exercées sur un atome i - c'est-à-dire une grandeur locale - nécessite de considérer l'intégralité des atomes de la boîte de simulation. Nous verrons plus loin (voir B.2) qu'en réalité, il est possible de ne se restreindre qu'au calcul d'énergie répulsive de l'atome et de ses voisins.

B.1 Dérivées d'un réseau de neurones

Déterminer des forces répulsives exercées sur un atome requiert, d'une part, d'évaluer la dérivée du réseau de neurones atomique par rapport à son vecteur d'entrée pour tous les atomes j du système.

$$\frac{\partial \mathrm{NN}\left(G(\mathbf{R}_{\mathbf{j}})\right)}{\partial G(\mathbf{R}_{\mathbf{j}})}$$

Nous adopterons la notation $\frac{\partial NN(\vec{X})}{\partial \vec{X}}$ à la fois plus légère et généralisable où l'on considère un vecteur d'entrée quelconque \vec{X} . De plus, notre application est un cas particulier dans le sens où les réseaux de neurones utilisés n'ont qu'un seul neurone de sortie; autrement dit, leur vecteur de sortie $\vec{Y} = NN(\vec{X})$ est de taille 1. Ce n'est évidemment pas toujours le cas dans d'autres applications. Aussi, dans cette section nous ferons l'effort de présenter le cas plus général où le vecteur de sortie \vec{Y} peut être de taille quelconque. Définissons les notations utilisées :

• \vec{X}^l est le vecteur d'entrée d'une couche l donnée. Il s'agit soit du vecteur de sortie de la couche l - 1 précédente, soit du vecteur d'entrée du réseau dans le cas de la couche d'entrée du réseau (l = 0).

- N_n^l est le nombre de neurones de la couche l; il s'agit donc de la taille du vecteur d'entrée de l + 1, \vec{X}^{l+1} .
- \vec{A}^{l} est le vecteur d'activation d'une couche l donnée. Pour les autres couches que la couche d'entrée $(l \neq 0)$, les éléments qu'il contient sont la somme pondérée des éléments du vecteur d'entrée \vec{X}^{l} . Pour un neurone donné n appartenant à l, on a donc :

$$A_{n}^{l} = \vec{w}_{n}^{l} \cdot \vec{X}^{l} + b_{n}^{l} = b_{n}^{l} + \sum_{m=1}^{N_{n}^{l-1}} w_{nm}^{l} \times X_{m}^{l}$$

où b_n^l est le biais du neurone et w_{nm}^l est le poids liant les neurones m de la couche l-1 et n de la couche l. Dans le cas de la couche d'entrée (l=0), on a simplement $\vec{A^0} = \vec{X^0}$.

• $\vec{Y}^l = \vec{X}^{l+1}$ est le vecteur de sortie de la couche *l*. Il s'agit de l'application de la fonction de transfert f^l de la couche *l* à l'activation $\vec{A^l}$.

$$Y_n^l = f^l \left(A_n^l \right)$$

Dans le cas de la couche d'entrée (l = 0), la fonction d'activation est linéaire, de la forme $f^0(x) = x$, et on considère donc que $\vec{Y}^0 = \vec{A}^0 = \vec{X}^0$.

Chaque couche l possède sa matrice de poids $\mathbf{w}^{\mathbf{l}}$ de taille $N_n^l \times N_n^{l-1}$ et son vecteur de biais \vec{b}^l de taille N_n^l . Pour un réseau à L couches cachées et de vecteur d'entrée \vec{X}^0 on détermine le vecteur de sortie $\vec{Y}^{L+1} = \mathrm{NN}(\vec{X}^0)$ comme :

$$\vec{Y}^{L+1} = f^{L+1} \left(\vec{A}^{L+1} \right) = f^{L+1} \left(\mathbf{w}^{\mathbf{L}+1} \cdot \vec{X}^{L+1} + b^{\vec{L}+1} \right) = f^{L+1} \left(\mathbf{w}^{\mathbf{L}+1} \cdot \vec{Y}^{L} + b^{\vec{L}+1} \right)$$

$$\vec{Y}^{L} = f^{L} \left(\vec{A}^{L} \right) = f^{L} \left(\mathbf{w}^{\mathbf{L}} \cdot \vec{X}^{L} + b^{\vec{L}} \right) = f^{L} \left(\mathbf{w}^{\mathbf{L}} \cdot \vec{Y}^{L-1} + b^{\vec{L}} \right)$$

$$\vdots$$

$$\vec{Y}^{1} = f^{1} \left(\vec{A}^{1} \right) = f^{1} \left(\mathbf{w}^{1} \cdot \vec{X}^{1} + b^{\vec{1}} \right) = f^{1} \left(\mathbf{w}^{1} \cdot \vec{Y}^{0} + b^{\vec{1}} \right) = f^{1} \left(\mathbf{w}^{1} \cdot \vec{X}^{0} + b^{\vec{1}} \right).$$

(B.5)

Il est donc possible de décrire la transformation T^l appliquée par une couche l du réseau ainsi :

$$T^{l}(\vec{X}) = f^{l}\left(A^{l}\left(\vec{X}\right)\right) = f^{l}\left(\mathbf{w}^{1} \cdot \vec{X} + \vec{b}^{l}\right).$$
(B.6)

La dérivée du réseau de neurones par rapport à son vecteur d'entrée $\frac{\partial \vec{Y}^{L+1}}{\partial \vec{X}^0}$ peut être écrite comme une matrice jacobienne associée à la transformation $NN(\vec{X}^0)$.

$$\begin{bmatrix} \frac{\partial Y_1^{L+1}}{\partial X_1^0} & \cdots & \frac{\partial Y_1^{L+1}}{\partial X_{N_n^0}^0} \\ \vdots & \ddots & \vdots \\ \frac{\partial Y_{N_n^{L+1}}^{L+1}}{\partial X_1^0} & \cdots & \frac{\partial Y_{N_n^{L+1}}^{L+1}}{\partial X_{N_n^0}^0} \end{bmatrix} = J(NN, \vec{X}^0).$$
(B.7)

Comme on peut le voir dans l'équation (B.5), la fonction NN peut être écrite comme une fonction composée de transformations successives T^l ; en appliquant le théorème de dérivation des fonctions composées à plusieurs variables, on peut réécrire le jacobien (B.7).

$$J\left(\mathrm{NN},\vec{X}^{0}\right) = J\left(T^{L+1},\vec{X}^{L+1}\right) \cdot J\left(T^{L},\vec{X}^{L}\right) \cdot \cdots \cdot J\left(T^{1},\vec{X}^{1}\right) \cdot J\left(T^{0},\vec{X}^{0}\right).$$
(B.8)

On peut donc réduire le problème de dérivation d'un réseau de neurones à la simple évaluation des jacobiens des transformations $F^l(\vec{X})$ prenant place dans chacune des couches. Ces transformations sont elles mêmes des fonctions composées (B.6), d'où :

$$\frac{\partial T^{l}}{\partial \vec{X}} = \frac{\partial f^{l}\left(\vec{A^{l}}\right)}{\partial \vec{A^{l}}} \cdot \frac{\partial \vec{A^{l}}}{\partial \vec{X}}.$$

Pour rappel, l'activation d'un réseau est la somme pondérée de ses valeurs d'entrées, sauf pour la couche d'entrée où il s'agit du vecteur d'entrée tel quel :

$$\vec{A}_n^l = \begin{cases} b_n^l + \sum_{m=1}^{N_n^{l-1}} w_{nm}^l \cdot X_m^l & \text{si } l \neq 0 \\ X_n^l & \text{si } l = 0 \end{cases} \rightarrow \frac{\partial \vec{A}_n^l}{\partial X_m^l} = \begin{cases} w_{nm}^l & \text{si } l \neq 0 \\ 1 & \text{si } l = 0 \text{ et } m = n \\ 0 & \text{sinon} \end{cases}.$$

Le terme restant de la dérivée, $\frac{\partial f^l(\vec{A}^l)}{\partial \vec{A}^l}$, est dépendant de la fonction de transfert utilisée par la couche du réseau de neurones.

• Fonction d'activation sigmoïde

$$f^{s}(x) = \frac{1}{1 + e^{-x}} \to \frac{\partial f^{s}(x)}{\partial x} = \frac{e^{-x}}{(1 + e^{-x})^{2}}.$$
 (B.9)

• Fonction d'activation tangente hyperbolique

$$f^t(x) = \tanh(x) \to \frac{\partial f^t(x)}{\partial x} = \left(1 - \tanh^2(x)\right).$$
 (B.10)

• Fonction d'activation linéaire

$$f^{l}(x) = x \to \frac{\partial f^{s}(x)}{\partial x} = 1.$$
 (B.11)

Au final, pour toute couche différente de la couche d'entrée $(l \neq 0)$, c'est-à-dire pour toute couche effectuant une transformation et ayant des poids et des biais, on peut écrire :

$$J\left(T^{l}, \vec{X}^{l}\right) = \begin{bmatrix} w_{1,1}^{l} \frac{\partial f^{l}\left(A_{1}^{l}\right)}{\partial A_{1}^{l}} & \cdots & w_{1,N_{n}^{l}}^{l} \frac{\partial f^{l}\left(A_{1}^{l}\right)}{\partial A_{1}^{l}} \\ \vdots & \ddots & \vdots \\ w_{N_{n}^{l-1},1}^{l} \frac{\partial f^{l}\left(A_{N_{n}^{l}}^{l}\right)}{\partial A_{N_{n}^{l}}^{l}} & \cdots & w_{N_{n}^{l-1},N_{n}^{l}}^{l} \frac{\partial f^{l}\left(A_{N_{n}^{l}}^{l}\right)}{\partial A_{N_{n}^{l}}^{l}}. \end{bmatrix}$$
(B.12)

B.2 Dérivées des fonctions de symétrie

Nous cherchons à déterminer l'expression analytique des forces répulsives (B.4),

$$F_{i,\alpha}^{\text{rep,NN}} = -\sum_{j=1}^{N_{\text{atoms}}^{\text{boite}}} \frac{\partial \text{NN} \left(G(\mathbf{R}_{j}) \right)}{\partial G(\mathbf{R}_{j})} \cdot \frac{\partial G(\mathbf{R}_{j})}{\partial r_{i,\alpha}} , \, \alpha = \left\{ x, y, z \right\},$$

et avons défini une démarche permettant d'évaluer le premier terme du produit précédemment (voir. B.1). Pour rappel (voir 5.4.1), les fonctions de symétrie centrées sur les atomes sont réparties en deux catégories. D'une part, les fonctions de symétrie radiales (5.12) sont constituées d'interactions de paires.

$$G_2^i = \sum_{j=1}^{N_{\text{vois}}^i} g_2(r_{ij}) = \sum_{j=1}^{N_{\text{vois}}^i} e^{-\eta(r_{ij}-r_s)^2} \cdot f_{\text{cut}}(r_{ij}).$$

D'autre part, les fonctions de symétrie angulaires (5.13) se composent d'interactions de triplets d'atomes.

$$\begin{aligned} G_4^i &= \sum_{j=1}^{N_{\text{vois}}^i} \sum_{\substack{k=1\\k\neq j}}^{N_{\text{vois}}^i} g_4\left(r_{ij}, r_{ik}, r_{jk}\right) \\ &= 2^{1-\zeta} \sum_{j=1}^{N_{\text{vois}}^i} \sum_{\substack{k=1\\k\neq j}}^{N_{\text{vois}}^i} (1+\lambda\cos\theta_{ijk})^{\zeta} \cdot e^{-\eta r_{ij}^2} \cdot e^{-\eta r_{ik}^2} \cdot e^{-\eta r_{jk}^2} \cdot f_{\text{cut}}(r_{ij}) \cdot f_{\text{cut}}(r_{ik}) \cdot f_{\text{cut}}(r_{jk}). \end{aligned}$$

Dans ces deux définitions, les fonctions de coupure utilisées sont les mêmes (5.11).

$$f_{\rm cut}(r) = \begin{cases} \tanh^3 \left(1 - \frac{r}{r_{\rm cut}}\right) & \text{si} \quad 0 < r < r_{\rm cut}, \\ 0 & \text{sinon} & . \end{cases}$$

Ces fonctions de symétrie sont appliquées à l'environnement atomique \mathbf{R}_i d'un atome i; elles sont dépendantes de la position $\vec{r_i}$ de l'atome central de l'environnement, mais aussi de celles de tous ses voisins $\{\vec{r_j}\}_{j=1}^{N_{\text{vois}}^i}$. On cherche à calculer $\frac{\partial G_i}{\partial r_{k,\alpha}}$, et on doit donc déterminer la dérivée par rapport à trois types d'atomes :

• $\vec{r}_k = \vec{r}_i$, où l'on dérive par rapport à la position de l'atome central de l'environnement. Dans ce cas, les dérivées de chacun des termes de la somme sont à évaluer et on a :

$$\frac{\partial G_2^i}{\partial r_{i,\alpha}} = \sum_{j=1}^{N_{\text{vois}}^i} \frac{\partial g_2(r_{ij})}{\partial r_{i,\alpha}} \quad \text{et} \quad \frac{\partial G_4^i}{\partial r_{i,\alpha}} = \sum_{j=1}^{N_{\text{vois}}^i} \sum_{\substack{k=1\\k\neq j}}^{N_{\text{vois}}^i} \frac{\partial g_4\left(r_{ij}, r_{ik}, r_{jk}\right)}{\partial r_{i,\alpha}}.$$

• $\vec{r}_k \in {\{\vec{r}_j\}}_{j=1}^{N_{\text{vois}}^i}$, c'est-à-dire dans le cas où k est un atome voisin de i. L'atome k contribue au calcul des fonctions de symétrie; seules les dérivées de ses contributions doivent être déterminées :

$$\vec{r}_k \in \{\vec{r}_j\}_{j=1}^{N_{\text{vois}}^i}$$
, $\frac{\partial G_2^i}{\partial r_{k,\alpha}} = \frac{\partial g_2(r_{ij})}{\partial r_{k,\alpha}}$ et $\frac{\partial G_4^i}{\partial r_{k,\alpha}} = \sum_{\substack{j=1\\j \neq k}}^{N_{\text{vois}}^i} \frac{\partial g_4(r_{ij}, r_{ik}, r_{jk})}{\partial r_{k,\alpha}}.$

• $\vec{r}_k \notin {\{\vec{r}_j\}}_{j=1}^{N_{\text{vois}}^i}$, auquel cas k n'appartient pas au voisinage de *i* et ne contribue pas au calcul des fonctions de symétrie; la dérivée est donc nulle.

$$\vec{r}_k \notin \{\vec{r}_j\}_{j=1}^{N_{\text{vois}}^i}$$
, $\frac{\partial G_2^i}{\partial r_{k,\alpha}} = \frac{\partial G_4^i}{\partial r_{k,\alpha}} = 0.$

B.2.1 Dérivées des fonctions de symétrie radiales

En rappelant que

$$g_2(r_{ij}) = e^{-\eta(r_{ij}-rs)^2} \cdot f_{\text{cut}}(r_{ij})$$

et avec la distance inter-atomique $r_{ij} = \sqrt{(r_{i,x} - r_{j,x})^2 + (r_{i,y} - r_{j,y})^2 + (r_{i,z} - r_{j,z})^2}$, on remarque qu'il est là encore possible d'appliquer le théorème de dérivation des fonctions composées.

$$\frac{\partial g_2\left(r_{ij}\right)}{\partial r_{k,\alpha}} = \frac{\partial g_2\left(r_{ij}\right)}{\partial r_{ij}} \frac{\partial r_{ij}}{\partial r_{k,\alpha}} \ , \ k = \left\{i, j\right\}.$$

Intéressons nous à la dérivée de la distance r_{ij} par rapport aux positions de i ou de j. On remarque que :

$$\frac{\partial r_{ij}}{\partial r_{i,\alpha}} = \frac{r_{i,\alpha} - r_{j,\alpha}}{r_{ij}} = -\frac{\partial r_{ij}}{\partial r_{j,\alpha}},\tag{B.13}$$

d'où $\frac{g_2(r_{ij})}{\partial r_{i,\alpha}} = -\frac{g_2(r_{ij})}{\partial r_{j,\alpha}}$: nous pouvons nous contenter de calculer $\frac{g_2(r_{ij})}{\partial r_{i,\alpha}}$. Il nous reste à évaluer la dérivée de la fonction de symétrie par rapport à la distance.

$$\frac{\partial g_2(r)}{\partial r} = \frac{\partial}{\partial r} \left[e^{-\eta (r-r_s)^2} \cdot f_{\rm cut}(r) \right] = \frac{\partial e^{-\eta (r-r_s)^2}}{\partial r} \cdot f_{\rm cut}(r) + e^{-\eta (r-r_s)^2} \cdot \frac{\partial f_{\rm cut}(r)}{\partial r}.$$

La dérivée de la partie exponentielle se calcule comme :

$$\frac{\partial e^{-\eta(r-r_s)^2}}{\partial r} = -2\eta \cdot (r-r_s) \cdot e^{-\eta(r-r_s)^2},$$

et la seconde dérivée dépend évidemment de la fonction de coupure choisie; nous nous plaçons dans le cas où l'on utilise une fonction de coupure de type tangente hyperbolique (5.11).

$$\frac{\partial f_{\rm cut}(r)}{\partial r} = \begin{cases} \frac{3}{r_{\rm cut}} \cdot \frac{\tanh^2 \left(1 - \frac{r}{r_{\rm cut}}\right)}{\cosh^2 \left(1 - \frac{r}{r_{\rm cut}}\right)} & \text{si} \quad 0 < r < r_{\rm cut}, \\ 0 & \text{sinon} \end{cases}$$
(B.14)

Nous avons donc le nécessaire pour exprimer les dérivées des fonctions de symétrie radiales.

$$\frac{\partial G_2^i}{\partial r_{i,\alpha}} = \sum_{j=1}^{N_{vois}^i} \frac{r_{i,\alpha} - r_{j,\alpha}}{r_{ij}} \cdot e^{-\eta(r_{ij} - r_s)^2} \left[\frac{\partial f_{\text{cut}}(r_{ij})}{\partial r_{ij}} - 2\eta \cdot (r_{ij} - r_s) \cdot f_{\text{cut}}(r_{ij}) \right] \\
\frac{\partial G_2^i}{\partial r_{k,\alpha}} = \frac{r_{k,\alpha} - r_{i,\alpha}}{r_{ik}} \cdot e^{-\eta(r_{ik} - r_s)^2} \left[\frac{\partial f_{\text{cut}}(r_{ik})}{\partial r_{ik}} - 2\eta \cdot (r_{ik} - r_s) \cdot f_{\text{cut}}(r_{ik}) \right].$$
(B.15)

B.2.2 Dérivées des fonctions de symétrie angulaires

Le calcul des dérivées des fonctions de symétrie angulaires est similaire au calcul effectué précédemment (voir. B.2.1). Le processus est cependant beaucoup plus long et volumineux.

On cherche à calculer la dérivée de $g_4(r_{ij}, r_{ik}, r_{jk})$. Ici, deux cas sont à considérer - le cas où l'on dérive par rapport à la position de l'atome central *i* et celui où l'on dérive par rapport à la position de l'un des voisins, *j* ou *k*. Ces deux cas sont relativement similaires; la seule différence réside dans la dérivée de la partie angulaire. Pour commencer, détaillons la dérivée par rapport à \vec{r}_i , la position de l'atome central.

$$\frac{\partial g_4(r_{ij}, r_{ik}, r_{jk})}{\partial r_{i,\alpha}} = 2^{1-\zeta} \cdot e^{-\eta r_{jk}^2} \cdot f_{\text{cut}}(r_{jk}) \\ \times \frac{\partial}{\partial r_{i,\alpha}} \left[(1 + \lambda \cos(\theta_{ijk}))^{\zeta} e^{-\eta (r_{ij}^2 + r_{ik}^2)} f_{\text{cut}}(r_{ij}) f_{\text{cut}}(r_{ik}) \right].$$

Afin d'identifier précisément ce que l'on doit calculer, séparons cette dérivée en quatre termes.

$$\frac{\partial}{\partial r_{i,\alpha}} \left[(1 + \lambda \cos(\theta_{ijk}))^{\zeta} e^{-\eta(r_{ij}^{2} + r_{ik}^{2})} f_{\text{cut}}(r_{ij}) f_{\text{cut}}(r_{ik}) \right] = \frac{\partial \left(1 + \lambda \cos(\theta_{ijk})\right)^{\zeta}}{\partial r_{i,\alpha}} e^{-\eta(r_{ij}^{2} + r_{ik}^{2})} f_{\text{cut}}(r_{ij}) f_{\text{cut}}(r_{ik}) + (1 + \lambda \cos(\theta_{ijk}))^{\zeta} \frac{\partial e^{-\eta(r_{ij}^{2} + r_{ik}^{2})}}{\partial r_{i,\alpha}} f_{\text{cut}}(r_{ij}) f_{\text{cut}}(r_{ik}) + (1 + \lambda \cos(\theta_{ijk}))^{\zeta} e^{-\eta(r_{ij}^{2} + r_{ik}^{2})} \frac{\partial f_{\text{cut}}(r_{ij})}{\partial r_{i,\alpha}} f_{\text{cut}}(r_{ik}) + (1 + \lambda \cos(\theta_{ijk}))^{\zeta} e^{-\eta(r_{ij}^{2} + r_{ik}^{2})} f_{\text{cut}}(r_{ij}) \frac{\partial f_{\text{cut}}(r_{ik})}{\partial r_{i,\alpha}}.$$
(B.16)

Où les termes à calculer sont :

• La partie angulaire des fonctions de symétrie g_4 .

$$\frac{\partial \left(1 + \lambda \cos(\theta_{ijk})\right)^{\zeta}}{\partial r_{i,\alpha}}$$

$$\frac{\partial \left(1 + \lambda \cos(\theta_{ijk})\right)^{\zeta}}{\partial r_{i,\alpha}} = \frac{\partial \left(1 + \lambda \cos(\theta_{ijk})\right)^{\zeta}}{\partial \theta_{ijk}} \frac{\theta_{ijk}(r_{i,\alpha})}{\partial r_{i,\alpha}}$$

Calculons tout d'abord

$$\frac{\partial \left(1 + \lambda \cos(\theta_{ijk})\right)^{\zeta}}{\partial \theta_{ijk}} = -\lambda \zeta \sin(\theta_{ijk}) (1 + \lambda \cos(\theta_{ijk}))^{\zeta - 1}.$$

Utilisons la loi des cosinus pour exprimer θ_{ijk} :

$$\theta_{ijk} = \arccos\left(\frac{r_{ij}^2(r_{i,\alpha}) + r_{ik}^2(r_{i,\alpha}) - r_{jk}^2}{2r_{ij}(r_{i,\alpha})r_{ik}(r_{i,\alpha})}\right) = \arccos(h(r_{i,\alpha})).$$

où $h(r_{i,\alpha}) = \frac{r_{ij}^2(r_{i,\alpha}) + r_{ik}^2(r_{i,\alpha}) - r_{jk}^2}{2r_{ij}(r_{i,\alpha})r_{ik}(r_{i,\alpha})}$. On peut donc calculer

$$\frac{\partial \theta_{ijk}(r_{i,\alpha})}{\partial r_{i,\alpha}} = \frac{\partial \arccos\left(h(r_{i,\alpha})\right)}{\partial r_{i,\alpha}} = \frac{\partial \arccos(h)}{\partial h} \frac{\partial h(r_{i,\alpha})}{\partial r_{i,\alpha}} = \frac{-1}{\sqrt{(1-h^2)}} \frac{\partial h(r_{i,\alpha})}{\partial r_{i,\alpha}}$$

Dérivons maintenant $h(r_{i,\alpha}) = \frac{r_{ij}^2(r_{i,\alpha}) + r_{ik}^2(r_{i,\alpha}) - r_{jk}^2}{2r_{ij}(r_{i,\alpha})r_{ik}(r_{i,\alpha})} = \frac{h^{\text{haut}}(r_{i,\alpha})}{h^{\text{bas}}(r_{i,\alpha})}$ $\frac{\partial \frac{h^{\text{haut}}(r_{i,\alpha})}{h^{\text{bas}}(r_{i,\alpha})}}{\partial r_{i,\alpha}} = \frac{\frac{\partial h^{\text{haut}}(r_{i,\alpha})}{\partial r_{i,\alpha}}h^{\text{bas}}(r_{i,\alpha}) - h^{\text{haut}}(r_{i,\alpha})\frac{\partial h^{\text{bas}}(r_{i,\alpha})}{\partial r_{i,\alpha}}}{(h^{\text{bas}}(r_{i,\alpha}))^2}.$

Afin d'obtenir la dérivée de $h^{\text{haut}}(r_{i,\alpha}) = r_{ij}^2(r_{i,\alpha}) + r_{ik}^2(r_{i,\alpha}) - r_{jk}^2$, rappelons (B.13) que $\frac{\partial r_{ij}(r_{i,\alpha})}{\partial r_{i,\alpha}} = \frac{r_{i,\alpha} - r_{j,\alpha}}{r_{ij}}$, d'où $\frac{\partial r_{ij}^2(r_{i,\alpha})}{\partial r_{i,\alpha}} = 2(r_{i,\alpha} - r_{j,\alpha})$, et $\frac{\partial h^{\text{haut}}(r_{i,\alpha})}{\partial r_{i,\alpha}} = \frac{\partial (r_{ij}^2(r_{i,\alpha}) + r_{ik}^2(r_{i,\alpha}) - r_{jk}^2)}{\partial r_{i,\alpha}} = 4r_{i,\alpha} - 2r_{j,\alpha} - 2r_{k,\alpha}.$

Ensuite, la dérivée de $h^{\text{bas}}(r_{i,\alpha}) = 2r_{ij}(r_{i,\alpha})r_{ik}(r_{i,\alpha})$ est :

$$\frac{\partial h^{\text{bas}}(r_{i,\alpha})}{\partial r_{i,\alpha}} = \frac{\partial (2r_{ij}(r_{i,\alpha})r_{ik}(r_{i,\alpha}))}{\partial r_{i,\alpha}} = 2\left[\frac{r_{ik}}{r_{ij}}(r_{i,\alpha} - r_{j,\alpha}) + \frac{r_{ij}}{r_{ik}}(r_{i,\alpha} - r_{k,\alpha})\right].$$

Nous pouvons désormais réécrire

$$\frac{\frac{\partial h(r_{i,\alpha})}{\partial r_{i,\alpha}}}{2r_{ij}r_{ik}\left(2r_{i,\alpha}-r_{j,\alpha}-r_{k,\alpha}\right)-\left(r_{ij}^{2}+r_{ik}^{2}-r_{jk}^{2}\right)\left[\frac{r_{ik}}{r_{ij}}(r_{i,\alpha}-r_{j,\alpha})+\frac{r_{ij}}{r_{ik}}(r_{i,\alpha}-r_{k,\alpha})\right]}{2r_{ij}^{2}r_{ik}^{2}}$$

On peut finalement exprimer la dérivée de θ_{ijk} :

$$\begin{aligned} \frac{\partial \theta_{ijk}}{\partial r_{i,\alpha}} &= \frac{-1}{\sqrt{1 - \frac{(r_{ij}^2 + r_{ik}^2 - r_{jk}^2)^2}{4r_{ij}^2 r_{ik}^2}}} \\ &\times \frac{2r_{ij}r_{ik}\left(2r_{i,\alpha} - r_{j,\alpha} - r_{k,\alpha}\right) - \left(r_{ij}^2 + r_{ik}^2 - r_{jk}^2\right)\left[\frac{r_{ik}}{r_{ij}}(r_{i,\alpha} - r_{j,\alpha}) + \frac{r_{ij}}{r_{ik}}(r_{i,\alpha} - r_{k,\alpha})\right]}{2r_{ij}^2 r_{ik}^2}. \end{aligned}$$

Nous pouvons enfin exprimer la dérivée analytique de la partie angulaire des fonctions de symétrie g_4 que nous cherchons à calculer.

$$\frac{\partial \left(1 + \lambda \cos(\theta_{ijk})\right)^{\zeta}}{\partial r_{i,\alpha}} = -\lambda \zeta \sin(\theta_{ijk}) (1 + \lambda \cos(\theta_{ijk}))^{\zeta - 1} \frac{-1}{\sqrt{1 - \frac{(r_{ij}^2 + r_{ik}^2 - r_{jk}^2)^2}{4r_{ij}^2 r_{ik}^2}}} \cdot \frac{2r_{ij}r_{ik} \left(2r_{i,\alpha} - r_{j,\alpha} - r_{k,\alpha}\right) - (r_{ij}^2 + r_{ik}^2 - r_{jk}^2) \left[\frac{r_{ik}}{r_{ij}} (r_{i,\alpha} - r_{j,\alpha}) + \frac{r_{ij}}{r_{ik}} (r_{i,\alpha} - r_{k,\alpha})\right]}{2r_{ij}^2 r_{ik}^2}.$$
(B.17)

• La partie exponentielle des fonctions de symétrie g_4 .

$$\frac{\partial e^{-\eta(r_{ij}^2+r_{ik}^2)}}{\partial r_{i,\alpha}} = \frac{\partial e^{-\eta r_{ij}^2}}{\partial r_{i,\alpha}} e^{-\eta r_{ik}^2} + e^{-\eta r_{ij}^2} \frac{\partial e^{-\eta r_{ik}^2}}{\partial r_{i,\alpha}}.$$

 $\frac{\partial e^{-\eta(r_{ij}^2+r_{ik}^2)}}{\partial r_{i,\alpha}}$

Cela requiert

$$\frac{\partial e^{-\eta r_{ij}^2(r_{i,\alpha})}}{\partial r_{i,\alpha}} = \frac{\partial e^{-\eta r_{ij}^2}}{\partial r_{ij}} \frac{\partial r_{ij}(r_{i,\alpha})}{\partial r_{i,\alpha}} = -2\eta e^{-\eta r_{ij}^2} \times (r_{i,\alpha} - r_{j,\alpha}).$$

On a donc

$$\frac{\partial e^{-\eta(r_{ij}^2 + r_{ik}^2)}}{\partial r_{i,\alpha}} = -2\eta(r_{i,\alpha} - r_{j,\alpha})e^{-\eta r_{ij}^2}e^{-\eta r_{ik}^2} - 2\eta e^{-\eta r_{ij}^2}(r_{i,\alpha} - r_{k,\alpha})e^{-\eta r_{ik}^2}$$

$$= -2\eta e^{-\eta(r_{ij}^2 + r_{ik}^2)}\left(2r_{i,\alpha} - r_{j,\alpha} - r_{k,\alpha}\right).$$
(B.18)

• $\frac{\partial f_{\text{cut}}(r_{ij})}{\partial r_{i,\alpha}} \rightarrow \text{calculé avec l'équation (B.14)}.$ • $\frac{\partial f_{\text{cut}}(r_{ik})}{\partial r_{i,\alpha}} \rightarrow \text{calculé avec l'équation (B.14)}.$

Une fois tous ces termes calculés explicitement, on peut réécrire l'équation B.16 pour exprimer analytiquement les dérivées des fonctions g_4 .

$$\frac{\partial g_4(r_{ij}, r_{ik}, r_{jk})}{\partial r_{i,\alpha}} = g_4(r_{ij}, r_{ik}, r_{jk}) \times \left[\frac{-\lambda \zeta \sin(\theta_{ijk})}{(1 + \lambda \cos(\theta_{ijk})} \frac{-1}{\sqrt{1 - \frac{(r_{ij}^2 + r_{ik}^2 - r_{jk}^2)^2}{4r_{ij}^2 r_{ik}^2}}} \right] \\
\frac{2r_{ij}r_{ik} (2r_{i,\alpha} - r_{j,\alpha} - r_{k,\alpha}) - (r_{ij}^2 + r_{ik}^2 - r_{jk}^2) \left[\frac{r_{ik}}{r_{ij}} (r_{i,\alpha} - r_{j,\alpha}) + \frac{r_{ij}}{r_{ik}} (r_{i,\alpha} - r_{k,\alpha}) \right]}{2r_{ij}^2 r_{ik}^2} \\
-2\eta (2r_{i,\alpha} - r_{j,\alpha} - r_{k,\alpha}) + \frac{r_{i,\alpha} - r_{j,\alpha}}{r_{ij}} \frac{\frac{\partial f_{\text{cut}}(r_{ij})}{\partial r_{ij}}}{f_{\text{cut}}(r_{ij})} + \frac{r_{i,\alpha} - r_{k,\alpha}}{r_{ik}} \frac{\frac{\partial f_{\text{cut}}(r_{ik})}{\partial r_{ik}}}{f_{\text{cut}}(r_{ik})} \right].$$
(B.19)

_

Dans le cas où l'on s'intéresse non plus à la dérivée par rapport à la position de l'atome central i mais par rapport à la position de l'un des voisins j (ou k, similairement), $\frac{\partial g^4(r_{ij},r_{ik})}{\partial r_{j,\alpha}}$, quelques modifications sont de mise :

- pour $\frac{\partial f_{\text{cut}}(r_{ij})}{\partial r_{j,\alpha}}$, $\frac{\partial f_{\text{cut}}(r_{jk})}{\partial r_{j,\alpha}}$ et $\frac{\partial e^{-\eta(r_{ij}^2 + r_{jk}^2)}}{\partial r_{j,\alpha}}$ rien ne change. Dans la dérivation de $\frac{\partial (1 + \lambda \cos(\theta_{ijk}))^{\zeta}}{\partial r_{j,\alpha}}$, il nous faut évaluer la dépendance en $r_{j,\alpha}$ de h.

$$h(r_{j,\alpha}) = \frac{r_{ij}^2(r_{j,\alpha}) + r_{ik}^2 - r_{jk}^2(r_{j,\alpha})}{2r_{ij}(r_{j,\alpha})r_{ik}} = \frac{h^{\text{haut}}(r_{j,\alpha})}{h^{\text{bas}}(r_{j,\alpha})}.$$

On peut réécrire :

$$\frac{\partial h^{\text{haut}}(r_{j,\alpha})}{\partial r_{j,\alpha}} = 2(r_{k,\alpha} - r_{i,\alpha}) \text{ au lieu de } 4r_{j,\alpha} - 2r_{i,\alpha} - 2r_{k,\alpha},$$

 et

$$\frac{\partial h^{\mathrm{bas}}(r_{j,\alpha})}{\partial r_{j,\alpha}} = 2\frac{r_{ik}}{r_{ij}}(r_{j,\alpha} - r_{i,\alpha}) \text{ au lieu de } 2\left[\frac{r_{jk}}{r_{ij}}(r_{j,\alpha} - r_{i,\alpha}) + \frac{r_{ij}}{r_{jk}}(r_{j,\alpha} - r_{k,\alpha})\right].$$

La dérivée de h est désormais :

$$\frac{h(r_{j,\alpha})}{\partial r_{j,\alpha}} = \frac{2r_{ij}r_{ik}(r_{k,\alpha} - r_{i,\alpha}) - \frac{r_{ik}}{r_{ij}}(r_{ij}^2 + r_{ik}^2 - r_{jk}^2)(r_{j,\alpha} - r_{i,\alpha})}{r_{ij}^2r_{ik}^2}.$$

On peut donc réécrire la dérivée finale en fonction de la position de j.

$$\frac{\partial g_4(r_{ij}, r_{ik}, r_{jk})}{\partial r_{j,\alpha}} = g_4(r_{ij}, r_{ik}, r_{jk}) \times \left[\frac{-\lambda \zeta \sin(\theta_{ijk})}{1 + \lambda \cos(\theta_{ijk})} \frac{-1}{\sqrt{1 - \frac{(r_{ij}^2 + r_{ik}^2 - r_{jk}^2)^2}{4r_{ij}^2 r_{ik}^2}}} \right] \cdot \frac{2r_{ij}r_{ik}(r_{k,\alpha} - r_{i,\alpha}) - \frac{r_{ik}}{r_{ij}}(r_{ij}^2 + r_{ik}^2 - r_{jk}^2)(r_{j,\alpha} - r_{i,\alpha})}{r_{ij}^2 r_{ik}^2} - 2\eta (2r_{j,\alpha} - r_{i,\alpha} - r_{k,\alpha}) + \frac{r_{j,\alpha} - r_{i,\alpha}}{r_{ij}} \frac{\partial f_{\text{cut}}(r_{ij})}{\partial r_{ij}} + \frac{r_{j,\alpha} - r_{k,\alpha}}{r_{jk}} \frac{\partial f_{\text{cut}}(r_{jk})}{\partial r_{ik}} \right].$$
(B.20)

Annexe C

Inclusion des outils d'apprentissage automatique dans DFTB+

L'idée ici est d'expliquer en détail la jonction entre la méthode DFTB et les outils d'apprentissage automatique, des choix techniques de l'interfaçage à l'évaluation de ses performances. Ce chapitre se construit suivant ces deux directions : d'une part, une présentation succincte de DFTB+, de la greffe en son sein des outils d'apprentissage automatique et des spécificités associées, et d'autre part une étude comparative du coût numérique par rapport aux anciennes implémentations des potentiels répulsifs.

C.1 Intégration dans DFTB+

Nous avons vu dans le chapitre 3 que l'énergie répulsive du système est historiquement calculée comme une somme de contributions de paires. En conséquence, la nouvelle représentation de ce terme comme une somme de contributions atomiques à plusieurs corps implique un changement conceptuel notable, que l'on rencontre inévitablement en se confrontant au code.

Cette section s'attache, dans un premier temps, à présenter la structure globale de DFTB+ et plus particulièrement le détail du calcul des interactions répulsives. Par la suite, on propose de présenter une librairie *Fortran*, développée au cours de cette thèse et permettant d'utiliser des réseaux de neurones et des fonctions de symétrie dans une logique fortement orientée vers la simulation atomistique.

C.1.1 Fonctionnement de DFTB+

La structure de DFTB+ permet en principe une forte évolutivité et facilite l'intégration de nos outils : le code est très granulaire et les interdépendances sont moindres. Dès la prise en main du code, on identifie immédiatement les parties dédiées, entre autres, aux solveurs des parties électroniques, aux différents outils de dynamique moléculaire ou d'optimisation de géométrie ou encore la gestion des entrées et sorties. L'identification des ajouts requis pour une nouvelle implémentation est directe; les régions du code à modifier sont minimales, réduites aux strictes contraintes inhérentes au modèle.

Evidemment, c'est également le cas pour le traitement des interactions répulsives. On se propose de présenter spécifiquement deux points, à savoir la structure de leur implémentation d'une part, et d'autre part comment on les utilise dans DFTB+.

Dans le code, le traitement des potentiels répulsifs est construit autour de deux types d'interactions de paires : les types repPoly, traitant des potentiels répulsifs de type polynomiaux, et repSplines, permettant d'utiliser des potentiels répulsifs construits par morceaux¹.

Les différents types d'interactions répulsives sont cachées au reste du code, qui ne voit qu'un objet plus global, de type repCont et qui tient lieu de conteneur. Ce choix de construction est extrêmement pertinent, notamment dans le cas où plusieurs éléments chimiques sont inclus dans la boîte de simulation : il est possible, voir même inévitable, que les différents éléments chimiques interagissent différemment entre eux. Non seulement les paramètres des différents polynômes et splines ne sont pas identiques selon que l'on considère par exemple une liaison Si-Si ou Si-O, mais il est également possible que l'un des potentiels soit polynomial et que l'autre soit de type splines. Ainsi, au lieu de s'embarrasser à manipuler un nombre indéfini d'instances de repPoly et repSplines, on les range dans un conteneur repCont, aisé à manipuler pour le reste du code.

Enfin, il est nécessaire de faire le lien entre les autres parties du code, qui fournissent des positions atomiques et qui récupèrent en sortie les énergies et forces répulsives, et les types repPoly et repSplines contenus dans repCont, qui ne manipulent que des distances interatomiques. Ce lien est fait par un jeu de routines, contenues dans le bien-nommé «*repulsives.F90*». L'idée est simple : à partir des positions des atomes de la boîte (et de leurs éventuelles images périodiques), on isole les paires d'atomes à transmettre au conteneur repCont et on récupère et retourne les énergies et forces répulsives.

Celles-ci sont seulement convoquées en quelques endroits spécifiques de DFTB+, dont les plus évidents sont les appels au calcul de l'énergie de boîte, appelés à chaque fois que l'on traite une géométrie donnée du système. Le calcul des forces répulsives est lui sollicité lorsque... la méthode requiert de calculer les forces. Cette évidence à part, les dérivées de l'énergie répulsive sont également nécessaires pour détermines les contraintes subies par le système via le calcul du *stress*.

C.1.2 Vue générale de la librairie et interfaçage

Le code DFTB+ étant écrit en *fortran*, nous avons choisi de développer une librairie sur-mesure, reposant sur les normes orientées objet Fortran 2003 et Fortran 2008. Cellesci implémentent les outils conceptuels nécessaires à notre application, que l'on présente ci-après (voir C.1.2.1).

^{1.} L'anglicisme *splines* s'est imposé dans l'usage en mathématiques pour désigner les fonctions définies par plusieurs polynômes appliqués chacun sur des morceaux d'intervalles.

Ce choix de développement, plutôt que l'emploi d'outils généraux (type *Tensorflow*) ou spécifiques (comme N2P2) se justifie de lui-même : au-delà de l'aspect didactique évident d'un tel exercice, il s'agit surtout la promesse d'un interfaçage aisé, propre et efficace à DFTB+.

Cependant, il est à noter que l'unique mode de fonctionnement requis pour cette librairie est le calcul de l'énergie répulsive (et des forces répulsives) d'une boîte de simulation avec des réseaux de neurones. Autrement dit, il n'est pas nécessaire que la librairie soit en mesure d'optimiser un réseau de neurones, puisque ce processus peut s'effectuer de manière complètement indépendante de DFTB+. Ainsi, et si la librairie s'enorgueillit tout de même de l'algorithme de rétropropagation du gradient appliqué aux énergies répulsives de boîtes et aux forces répulsives (voir 6.2.1), on délègue en pratique l'étape d'optimisation à un logiciel externe spécialisé - N2P2 - qui utilise des algorithmes d'optimisation plus raffinés et efficaces (voir 6.2.2).

C.1.2.1 Structure du code

En cherchant à rendre compte de la méthodologie globale présentée dans la trame principale des présents travaux, la librairie se construit autour de trois idées centrales.

D'une part, il est évidemment capital de manipuler aisément les notions relatives aux positions atomiques, point de départ de toute application. Ainsi, la définition d'un *environnement atomique* est l'une des clés de voûte de notre modèle; de nombreux outils et concepts s'y rattachent, comme par exemple la *boîte de simulation*, qui est une *composition* d'environnements.

D'autre part, les observables de sortie, c'est-à-dire $E_{\rm rep}$ et $F_{\rm rep}$, sont le fruit du calcul effectué par un *réseau de neurones*. Là encore, le traitement des données gérées par le réseau, mais aussi les *couches* et *neurones* le constituant sont directement liés à ce concept.

Enfin, la troisième pierre angulaire du modèle, le concept de descripteur (et particulièrement de *fonctions de symétrie centrées sur les atomes - acsf*), lie les deux précédentes. Cet objet, collection de jeux de paramètres, permet de manipuler les positions atomiques des environnements, et de produire les descripteurs et leurs dérivées requis par les réseaux de neurones.

Un diagramme UML^2 de la librairie est disponible plus loin annexe D) - on ne montre pas la plupart des routines privées des différentes classes par souci de lisibilité. La structure se veut épurée, et la construction est relativement explicite : une boîte est composée d'environnements, un réseau de neurones est composé de couches, elles-même composées de briques élémentaires, les neurones. Le principal apport de la programmation orientée objet réside d'ailleurs du côté des couches (*Layer*) des réseaux, et plus particulièrement sur la possible spécialisation des instances de ces classes dans deux classes filles : les couches d'entrées (*Input_layer*) et de sortie (*Output_layer*). Dans cet exemple de cas d'utilisa-

^{2.} Unified Modeling Language, ou Langage de Modélisation Unifié. Formalisme de modélisation graphique pour les applications orientées objet, permettant une visualisation conceptuelle affranchie des contraintes du langage utilisé.

tion de l'héritage, le fonctionnement pratique des deux classes filles est très proche de la classe mère, mais leurs différences conceptuelles sont assurées par construction.

Par ailleurs, et dans un objectif d'évolutivité maximale, il semble souhaitable de créer une classe abstraite *Descriptor* dont *acsf* hériterait, et qui guiderait l'éventuelle création d'une nouvelle classe fille implémentant une alternative aux calculs de descripteur. En pratique, l'absence de ce filet de sûreté, initialement présent, est assez peu dommageable : les calculs de descripteurs sont uniquement voués à retourner un vecteur de réels, et aucune instance de ces classes n'est manipulée par d'autres objets du code. L'ajout d'une classe abstraite ajoute une sécurité conceptuelle certaine, mais on peut aisément nuancer l'impérativité d'un tel choix. On peut en effet arguer que, au-delà de complexifier la compréhension sémantique du code pour un non initié au formalisme objet, la présence d'une classe abstraite imposerait un squelette aux futures implémentations qui serait au mieux superflu, et au pire contre productif car restreignant trop les futurs ajouts. Un exemple pertinent pour s'en convaincre est la méthode des matrices de Coulomb (voir 5.4.2.2), qui ne concentre plus sa représentation autour d'environnements atomiques mais d'environnements de liaisons et qui présente de nombreuses contraintes inhérentes spécifiques, telles qu'entre autres la gestion de la taille de son vecteur de sortie. Dans un tel cas, la structure imposée par une éventuelle classe abstraite Descriptor dont acsf et les matrices de Coulomb hériteraient serait absurdement simpliste (retourner uniquement un vecteur de réels) et donc plutôt inutile.

Globalement, la logique de la librairie est simple : la manipulation des entrées et sorties peut se faire sur disque. Ainsi, on peut lire (et si besoin écrire) des environnements atomiques, des boîtes de simulation, des paramètres de fonctions de symétrie et des réseaux de neurones dans des fichiers textes formatés; on manipule ensuite librement des instances de ces classes. À l'inverse, il est tout à fait envisageable de transmettre les informations importantes (positions atomiques, paramètres de descripteurs ou de réseau de neurones) par un code utilisant la librairie (tel que DFTB+ par exemple). En elle-même, la librairie n'a que deux principales applications : vérifier la précision d'un réseau de neurones sur un jeu de données (processus de calcul), et optimiser un réseau avec l'algorithme de rétropropagation sur un jeu de données (processus d'apprentissage, inutilisé dans DFTB+).

C.1.2.2 Normalisation

Avant de s'intéresser à l'interfaçage de la librairie au code DFTB+, attardons nous sur les différentes normalisations opérées afin de faciliter le calcul. Il s'agit de celles opérées par N2P2 en interne; elles sont implémentées au sein de la librairie car leur pertinence est analogue dans le cas où l'on veut reproduire des potentiels répulsifs.

Tout d'abord, on veut assurer que l'écart type des énergies normalisées soit égal à 1. Pour cela, on divise toutes les énergies par l'écart-type calculé sur le jeu de données de référence.

$$\sigma_E = \sqrt{\frac{1}{N_{\text{boites}} - 1} \sum_{i=1}^{N_{\text{boites}}} \left(\frac{E_i}{N_{\text{atoms}}^i} - \langle E \rangle\right)^2}.$$

On note ici $\langle E \rangle$ l'énergie moyenne par atome :

$$\langle E \rangle = \frac{1}{N_{\text{boites}}} \sum_{i=1}^{N_{\text{boites}}} \frac{E_i}{N_{\text{atomes}}^i}.$$

Au passage, remarquons que l'on calcule l'écart-type σ_E en appliquant la correction de Bessel : on divise la somme par n-1 plutôt que par n. Cette correction est fréquemment utilisée dans le cas où l'on considère un échantillon de la population totale que l'on souhaite représenter. Elle assure que la moyenne des écarts-types calculés sur tous les échantillons possibles soit égale à l'écart-type calculé sur la population totale. Ce choix semble pertinent dans notre cas, où les énergies présentes dans le jeu de données d'apprentissage ne constituent évidemment pas l'intégralité des énergies possibles pour les systèmes que l'on veut représenter.

Par ailleurs, on translate les énergies répulsives de référence de $\langle E \rangle$, pour que la moyenne des énergies translatées par atome présentes dans le jeu de référence soit nulle. On tient également compte de l'énergie d'un atome isolé $E_{\rm at}$ de manière à ce que les valeurs de l'énergie tendent à s'annuler au longues distances³. On peut donc associer $E_{\rm conv} = \frac{1}{\sigma_E}$ à un facteur de conversion en énergies, et la normalisation s'écrit au final :

$$E_i^{\text{norm}} = E_i \times E_{\text{conv}} - N_{\text{atomes}}^i \cdot (\langle E \rangle + E_{\text{at}}).$$

En raison de la normalisation des énergies présentée précédemment, les forces sont modifiées et s'écrivent :

$$-\frac{\partial}{\partial r_{j,\alpha}} \left[E_i \times E_{\text{conv}} - N_{\text{atomes}}^i \cdot (\langle E \rangle + E_{\text{at}}) \right] = F_{j,\alpha} \times E_{\text{conv}}.$$

De la même manière, on cherche à assurer que l'écart type des forces soit égal à 1. Ici, on a :

$$\sigma_F = \sqrt{\frac{1}{3N_{\text{atomes}}^{\text{total}} - 1} \sum_{i=1}^{N_{\text{boites}}} \sum_{j=1}^{N_{\text{atomes}}} \sum_{\alpha = \{x, y, z\}} (F_{j,\alpha} - \langle F \rangle)^2}$$

$$N_{\text{boites}}$$

où $N_{\text{atomes}}^{\text{total}} = \sum_{i=1}^{N_{\text{boites}}} N_{\text{atomes}}^{i}$, et :

$$\langle F \rangle = \frac{1}{3N_{\text{atomes}}^{\text{total}}} \sum_{i=1}^{N_{\text{boites}}} \sum_{j=1}^{N_{\text{atomes}}^{i}} \sum_{\alpha=\{x,y,z\}} F_{j,\alpha}$$

^{3.} En principe, notre construction (voir 7.1.3) veut que l'énergie répulsive d'atomes isolés soit nulle; par souci de généralisation, nous laissons tout de même la possibilité de modifier ce paramètre.

On peut identifier $L_{\text{conv}} = \frac{\sigma_F}{\sigma_E}$ comme un facteur de conversion de longueurs, et on a donc au final :

$$F_{j,\alpha}^{\text{norm}} = F_{j,\alpha} \times \frac{E_{\text{conv}}}{L_{\text{conv}}}.$$

Ces choix de normalisation permettent donc de faciliter le travail de représentation du réseau de neurones par la modification de la distribution des valeurs cibles de référence. Conjointement, il semble très pertinent d'avoir une approche similaire pour les vecteurs d'entrée des réseaux; intéressons nous à la normalisations des descripteurs. Plusieurs stratégies sont possibles :

• on peut translater les descripteurs à l'intérieur d'un intervalle $[S_{min}, S_{max}]$.

$$G^{norm} = S_{min} + (S_{max} - S_{min}) \frac{G - G_{min}}{G_{max} - G_{min}}.$$

• on peut simplement s'assurer que la valeur moyenne des descripteurs soit nulle.

$$G^{norm} = G - \langle G \rangle$$
, où $\langle G \rangle = \frac{1}{N_{\text{ACSF}}} \sum_{i=1}^{N_{\text{ACSF}}} G_i$

• on peut chercher à prédéterminer la moyenne des descripteurs $\langle G^{norm} \rangle = S_{min}$, et à s'assurer qu'ils soient dans un intervalle fixe $[2S_{min} - S_{max}, S_{max}]$.

$$G^{norm} = S_{min} + (S_{max} - S_{min}) \frac{G - \langle G \rangle}{G_{max} - G_{min}}.$$

<u>Preuves</u> :

$$\langle G^{norm} \rangle = \frac{1}{N_{\text{ACSF}}} \sum_{i=1}^{N_{\text{ACSF}}} \left[S_{min} + (S_{max} - S_{min}) \frac{G_i - \langle G \rangle}{G_{max} - G_{min}} \right]$$
$$= S_{min} - \frac{S_{max} - S_{min}}{G_{max} - G_{min}} \times \left(\langle G \rangle - \frac{1}{N_{\text{ACSF}}} \sum_{i=1}^{N_{\text{ACSF}}} G_i \right) = S_{min}.$$

et, puisque $S_{max} > S_{min}$:

$$-1 \leq \frac{G - \langle G \rangle}{G_{max} - G_{min}} \leq 1$$
$$-(S_{max} - S_{min}) \leq (S_{max} - S_{min}) \cdot \frac{G - \langle G \rangle}{G_{max} - G_{min}} \leq (S_{max} - S_{min})$$
$$2S_{min} - S_{max} \leq G^{norm} \leq S_{max}.$$

• enfin, on peut souhaiter avoir un jeu de descripteurs distribués de manière à avoir un écart-type égal à $\sigma^{norm} = S_{max} - S_{min}$ et de moyenne $\langle G^{norm} \rangle = S_{min}$.

$$G^{norm} = S_{min} + (S_{max} - S_{min}) \cdot \frac{G - \langle G \rangle}{\sigma}.$$

Tout comme précédemment, l'écart-type σ utilise la correction de Bessel.

$$\sigma = \sqrt{\frac{1}{N_{\text{ACSF}} - 1}} \sum_{i=1}^{N_{\text{ACSF}}} (G_i - \langle G \rangle)^2.$$

<u>Preuves</u> :

$$\langle G^{norm} \rangle = \frac{1}{N_{\text{ACSF}}} \sum_{i=1}^{N_{\text{ACSF}}} \left[S_{min} + (S_{max} - S_{min}) \frac{G_i - \langle G \rangle}{\sigma} \right]$$
$$= S_{min} - \frac{S_{max} - S_{min}}{\sigma} \times \left(\langle G \rangle - \frac{1}{N_{\text{ACSF}}} \sum_{i=1}^{N_{\text{ACSF}}} G_i \right) = S_{min}.$$

ce qui nous permet de calculer l'écart-type

$$\sigma^{norm} = \sqrt{\frac{1}{N_{ACSF} - 1} \sum_{i=1}^{N_{ACSF}} \left[S_{min} + \frac{S_{max} - S_{min}}{\sigma} \left(G_i - \langle G \rangle \right) - \langle G^{norm} \rangle \right]^2} \\ = \sqrt{\frac{1}{N_{ACSF} - 1} \sum_{i=1}^{N_{ACSF}} \left[\frac{S_{max} - S_{min}}{\sigma} \left(G_i - \langle G \rangle \right) \right]^2} \\ = \frac{S_{max} - S_{min}}{\sigma} \times \sqrt{\frac{1}{N_{ACSF} - 1} \sum_{i=1}^{N_{ACSF}} \left(G_i - \langle G \rangle \right)^2} \\ = S_{max} - S_{min}.$$

Le choix parmi toutes ces approches est laissé à l'utilisateur (par le biais de mots-clés) et il fixe également la valeur des constantes S_{min} et S_{max} .

C.1.2.3 Interfaçage à DFTB+

Comme présenté précédemment (voir C.1.1), on peut complètement résumer les potentiels répulsifs historiquement intégrés à DFTB+ en deux types d'interactions de paires (*repPoly* et *repSplines*), accessibles indirectement aux autres parties du code (calculs d'énergies, de forces et de stress) qui appellent les routines de calculs contenues dans «*repulsive.F90*» en manipulant un conteneur (*repCont*). En substance, il s'agit des principaux éléments dont nous devons tenir compte pour greffer la librairie au code.

Ainsi, on crée un troisième type repNeural, à la structure très proche de repPoly et repSplines, et qui est amené à être manipulé de manière analogue par une instance du conteneur repCont. La différence majeure de ce nouveau type vis-à-vis de ses aînés est cependant notable : si jusque là on considérait uniquement des potentiels de paire, notre ajout demande de représenter des interactions à plusieurs corps, c'est-à-dire de prendre en compte de l'intégralité de l'environnement d'un atome.

Cette divergence sémantique constitue l'unique défi majeur de cet interfaçage. Le code DFTB+ propose en pratique un accès à une liste de voisins ⁴ très proche du concept d'environnement atomique tel qu'on l'a définit ; ceci-dit, cette liste de voisins est pensée pour des interactions de paires, et quelques ajustements sont de mise.

Premièrement, et de manière assez triviale, il faut s'assurer de fournir à la librairie une liste de voisins ne contenant que les interactions couvertes par ses outils : le potentiel

^{4.} Liste des index et des positions des atomes proches d'un atome spécifique, avec possible prise en compte des éventuelles images périodiques nécessaires.

répulsif sélectionné dépend du type d'éléments chimiques en jeu, mais cette sélection n'est immédiate que lorsque l'on ne considère que des paires d'atomes. Prenons par exemple le cas où l'on a deux éléments chimiques Si et O présents dans la boîte. Supposons que nous ne sommes en capacité de représenter que les interactions Si-Si avec nos outils, et les interactions Si-O ou O-O sont respectivement traitées avec un potentiel polynômial et de type splines. Il faut dans un premier temps ignorer tous les atomes d'oxygène des différents environnements pour déterminer les contributions Si-Si à plusieurs corps, puis retomber efficacement sur le fonctionnement par défaut de DFTB+ pour sélectionner le potentiel de paire approprié pour gérer les paires Si-O et O-O.

Par ailleurs, la liste de voisins manipulée par DFTB+ utilise intelligemment une propriété clé des interactions de paires : la symétrie du potentiel répulsif pour une paire d'atomes donnée. Comme on ne tient compte que des distances r_{AB} entre atomes, l'énergie $V_{\rm rep}(r_{AB})$ exercée par A sur B est la même que celle $V_{\rm rep}(r_{BA})$ exercée par B sur A. En conséquence, les listes de voisins sont construites pour ne tenir compte qu'une seule fois de chaque paire : si on a déjà la paire A-B, on n'a pas besoin de B-A. Cela s'effectue en n'incluant un atome à la liste de voisin d'un autre que si son index est supérieur. Mis en équation, cela équivaut à

$$E_{\rm rep} = \sum_{A=1}^{N_{\rm atomes}} \sum_{B>A}^{N_{\rm atomes}} V_{\rm rep} \left(r_{AB} \right).$$

Ce choix est extrêmement courant car il permet de réduire drastiquement le temps de calcul : d'un coût en $O(N^2)$, on passe à un coût $O(N \log N)$, où $N = N_{\text{atomes}}$. Cependant, il n'est plus possible d'effectuer cette simplification si le potentiel considéré n'est pas un potentiel de paire. Une autre modification à la liste de voisins est donc de mise, et nous force à reconstruire une liste complète assurant que si A appartient à la liste de B, alors l'inverse doit aussi être vrai. Ce processus est mérite une attention particulière dans le cas de systèmes périodiques, car si A est voisin d'une image périodique de B, il faut identifier l'image périodique de A voisine de B (Fig. C.1), et l'une de ces images ne sera pas présente dans la liste de voisins orientée vers les interactions de paires.

Une fois les listes de voisins convenablement construites, la création d'environnements atomique $\mathbf{R}_{\mathbf{i}}$ est immédiate. On est alors à même de calculer les vecteurs descripteurs G_i associés, utilisés par le réseau de neurones pour retourner les contributions atomiques $E_{\mathrm{rep}}^{i,\mathrm{NN}}$ à l'énergie répulsive.

Par définition (B.2), le réseau de neurones calcule la composante $\alpha = \{x, y, z\}$ de la force répulsive $F_{i,\alpha}^{\text{rep,NN}}$ appliquée sur l'atome *i* comme :

$$F_{i,\alpha}^{\rm rep,NN} = -\sum_{j=1}^{N_{\rm atomes}^{\rm boite}} \frac{\partial E_{\rm rep}^{j,\rm NN}}{\partial r_{i,\alpha}},$$

que l'on réécrit

$$F_{i,\alpha}^{\text{rep,NN}} = -\sum_{j=1}^{N_{\text{atomes}}^{\text{boite}}} \frac{\partial E_{\text{rep}}^{j,\text{NN}}}{\partial G_j} \cdot \frac{\partial G_j}{\partial r_{i,\alpha}},$$



FIGURE C.1 – Sélection d'images périodiques pour reconstruire la liste de voisins.

or, on sait que $\frac{\partial G_j}{\partial r_{i,\alpha}} \neq 0$ si $i \in \mathbf{Rj}$, c'est-à-dire si i = j ou $r_{ij} < r_{\text{cut}}$.

$$F_{i,\alpha}^{\mathrm{rep,NN}} = -\frac{\partial E_{\mathrm{rep}}^{i,\mathrm{NN}}}{\partial G_i} \cdot \frac{\partial G_i}{\partial r_{i,\alpha}} - \sum_{j=1}^{N_{\mathrm{voisins}}^i} \frac{\partial E_{\mathrm{rep}}^{j,\mathrm{NN}}}{\partial G_j} \cdot \frac{G_j}{\partial r_{i,\alpha}}$$

Ici, on remarque que la force exercée sur l'atome i ne dépend pas que de son propre environnement atomique \mathbf{R}_i , mais aussi de celui de tous ses voisins. En conséquence, lors du calcul par un réseau de neurones de l'énergie répulsive d'un atome, on retourne si besoin la contribution aux forces appliquées sur cet atome mais aussi sur tous ses voisins.

C.2 Utilisation et performance

Du point de vue de l'utilisateur, deux ajustements sont requis pour profiter des potentiels répulsifs numériques. Il faut dans un premier temps spécifier, dans le fichier d'entrée principal du code (*dftb_in.hsd*), pour quelles paires d'éléments chimiques on souhaite recourir à des polynômes, des réseaux de neurones ou, par défaut, des splines. Par exemple, pour un système qui contient les espèces chimiques Si et O, on dénombre trois types d'interactions : Si-Si, Si-O et O-O. Dans le cas où les potentiels répulsifs respectifs seraient de type réseau de neurones pour Si-Si, polynomial pour Si-O et splines pour O-O, le fichier d'entrée *dftb_in.hsd* contiendrait : # Quelles paires d'éléments utilisent des potentiels répulsifs de type réseau de neurones ?

```
NeuralRepulsive = {
    Si-Si = Yes
    # Si une paire d'élément n'est pas spécifiée, la valeur par défaut est 'No'.
}
# Quelles paires d'éléments utilisent des potentiels répulsifs de type polynômiaux ?
PolynomialRepulsive = {
    Si-Si = No # Superflu, car égal à la valeur par défaut.
    Si-O = Yes
    # Si une paire d'élément n'est pas spécifiée, la valeur par défaut est 'No'.
}
```

N'ayant pas été spécifiée, l'interaction répulsive O-O est supposée de type splines.

Evidemment, ces spécifications ne sont pas suffisantes, seulement utiles pour spécifier au code quelles interactions rechercher dans les fichiers de paramètres. Ceux-ci, que l'on appelle «fichiers de *Slater-Koster*», sont spécifiques à une paire d'espèces chimiques, sont tabulés et contiennent les valeurs nécessaires (Chap. 3) aux calculs *DFTB*. Cela inclut les paramétrisations des parties électroniques du calcul, mais aussi celles nous intéressant présentement : les paramètres des potentiels répulsifs. Dans le cas d'une interaction de paires, cette section des fichiers s'amorce par un mot-clé (*«Polynomial»* ou *«Splines»*), puis on trouve les paramètres spécifiques à la méthode : distances, rayons de coupure, nombre de splines, coefficients polynômiaux, ... Dans le cas potentiels répulsifs de type réseaux de neurones, le mot-clé initiant le bloc est *«Neural-Network»*. Ensuite, les informations relatives au potentiel sont présentes ; dans l'ordre :

- la forme du réseau.
- le nombre de fonctions de symétrie radiales et angulaires.
- les paramètres $(\eta, r_s, r_{\text{cut}} \text{ et } \eta, \zeta, \lambda, r_{\text{cut}})$ des fonctions de symétrie.
- les constantes de normalisation du réseau de neurones $(S_{min}, S_{max}, E_{at}, E_{conv}, < E >, L_{conv}).$
- les constantes de normalisation de chaque fonction de symétrie G_i ($G_i^{max}, G_i^{min}, < G_i >, \sigma_{G_i}$).
- les poids et biais du réseau.
- optionnellement, les paramètres $\{r_0, a, b\}$ du potentiel répulsif *ad-hoc* à courte portée (voir 7.1.5).

Dans les faits, il est assez peu fréquent de construire soi-même un fichier de *Slater-Koster*; en règle générale, le rôle de l'utilisateur se limite donc à spécifier l'interaction choisie dans *dftb_in.hsd*, et à fournir un fichier de *Slater-Koster* adapté, sans besoin d'en connaître les détails de fonctionnements internes.

En revanche, il est tout à fait dans son intérêt d'être au fait des performances du potentiel choisi. La précision et la transférabilité sont bien-sûr essentielles; celles-ci sont présentées plus en détail dans le chapitre 8 de la trame principale. Ici, nous nous attachons à étudier l'autre pendant capital de DFTB+: l'efficacité numérique. La méthode est par essence semi-empirique, aspire à s'approcher de la précision de la théorie de la fonction-nelle de la densité, et est avant tout motivée par la volonté de réduire drastiquement
le coût numérique. Cela passe principalement par le calcul préalable d'un bon nombre de paramètres, permettant de s'affranchir de la diagonalisation coûteuse du hamiltonien *DFT*. Le choix d'un potentiel répulsif empirique à la forme simple va dans ce sens, et les parties électroniques occupent une large majorité du temps de calcul. Malgré l'augmentation notable de complexité apportée par la considération d'un potentiel répulsif à plusieurs corps, il est capital de s'assurer que le coût ne s'en trouve pas impacté trop fortement, et que la partie répulsive des calculs reste négligeable par rapport aux autres.

Plusieurs paramètres semblent nécessiter une investigation : on peut penser à la forme du réseau de neurones utilisé - le nombre de neurones par couche et/ou le nombre de couches cachées - mais aussi au type de fonction de symétrie utilisées et au rayon de coupure $r_{\rm cut}$. La figure C.2 montre la performance relative de plusieurs réseaux de neurones par rapport aux interactions de paires de type *splines* pour des boîtes de simulation de tailles variables (entre 2 et 2048 atomes). Ces réseaux sont construits de manière à varier tous les paramètres présentés précédemment : on change le nombre de neurones par couche avec 10, 50 ou 100 neurones d'entrées et respectivement 5, 25 et 50 neurones par couche cachée. On s'intéresse à des réseaux avec une, deux ou trois couches cachées. La répartition des fonctions de symétrie est également variable : uniquement des fonctions G_2 , uniquement des fonctions G_4 ou un mélange 50% - 50% de ces deux types de fonctions. Enfin, on étudie l'impact du rayon de coupure avec des fonctions de symétrie dont le rayon de coupure $r_{\rm cut}$ vaut 3 Å ou 6 Å. Au total, on étudie donc les performances de 54 réseaux.



FIGURE C.2 – Comparaison des temps de calcul DFTB+ de différents réseaux de neurones par rapport aux interactions de paires.

Dans cette figure, on remarque immédiatement que le principal facteur influant le coût numérique est, de manière assez prévisible, le rayon de coupure : plus celui-ci est grand, plus le calcul est long. Notons qu'ici, on compare des réseaux de neurones dont le rayon de coupure vaut respectivement $r_{\rm cut} = 3$ Å ou $r_{\rm cut} = 6$ Å avec un potentiel de paire de rayon de coupure $r_{\rm cut} = 2.54$ Å; un écart est donc attendu. On y remarque également que les réseaux avec un plus grand nombre de fonctions de symétrie présentent évidemment un coût numérique plus élevé, et surtout que le calcul des fonctions G_4 est sensiblement plus long. En revanche, il semble intéressant de souligner que ces différences de temps de calcul sont marginales dans le cas d'un rayon de coupure réduit. Cela s'explique aisément par le fait que le nombre de voisins croît exponentiellement avec l'augmentation du rayon de coupure, et l'écart en coût numérique sensible entre des fonctions G_2 et G_4 est donc bien plus perceptible. Enfin, il semble apparent que le nombre de couches cachées du réseau n'est pas significativement impactant quant à la performance du réseau.

Afin d'avoir une meilleure idée du comportement des réseaux pour des systèmes de taille plus modeste, portons notre attention à la figure C.3. On y représente les mêmes données que précédemment, mais l'on se contente d'observer les systèmes d'au maximum 128 atomes.



FIGURE C.3 – Comparaison des temps de calcul de différents réseaux de neurones par rapport aux interactions de paires : systèmes plus modestes.

Même lorsque l'on se concentre sur des systèmes plus petits, il est évident que les réseaux avec un grand rayon de coupure employant des fonctions de symétrie angulaires semblent significativement plus coûteux numériquement. D'autre part, on remarque de manière générale une certaine constance dans les performances numériques des réseaux dont le rayon de coupure est $r_{\rm cut} = 3$ Å. On note bien, ça et là, des fluctuations sensibles - notamment vers 50 atomes - pouvant soutenir la thèse d'un potentiel moins coûteux que les interactions de paires ; il semble plus prudent d'assumer que ces pics ponctuels sont le fruit de coïncidences numériques dues aux optimisations performées par le compilateur. En revanche, et au vu des deux figures précédentes, il paraît raisonnable d'affirmer que l'utilisation de potentiels numériques à plusieurs corps à courte portée sous la forme de réseaux de neurones est équivalent, en terme de coût de calcul, à l'emploi de potentiels répulsifs de paire.

On rappelle à titre indicatif que les réseaux utilisés dans les présents travaux ont deux couches cachées, utilisent au plus une cinquantaine de fonctions de symétrie - indifféremment des fonctions radiales G_2 ou angulaires G_4 - et ont nécessairement un rayon de coupure réduit ($r_{\rm cut} = 3$ Å). Au vu des résultats présentés, il est donc attendu que le coût de calcul soit analogue à celui des potentiels répulsifs de type *splines*. C'est en pratique ce que l'on observe au cours de l'intégralité des calculs présentés dans le reste de cette thèse.

Annexe D

Diagramme UML de la librairie



Annexe E

Résultats supplémentaires : approches purement numériques

Ce chapitre présente des propriétés obtenues via deux potentiels purement numériques, construits pour représenter directement les énergies totales et les forces interatomiques issues de calculs DFT. Le premier d'entre eux, nommé **NNP**, est obtenu d'après les données de [98] pour le Silicium pur, tandis que le second, nommé **GAP**, est celui proposé dans [74]. Ces deux outils sont présentés plus en détail dans la trame principale du présent manuscrit (Chap. 8), et de nombreux résultats y sont également inclus.

Certaines figures se prêtent assez mal à la présentation conjointe d'un trop grand nombre de résultats; pour favoriser la lisibilité, on choisit dans ces cas précis de présenter ces figures dans la présente section annexe.

Commençons par l'étude de la stabilité relative des cristaux en fonction du volume pour ces deux approches purement paramétriques. On les compare dans la figure E.1 à celles prédites par la méthode premier principe de référence, VASP.

Ces résultats sont discutés plus extensivement et remis en perspective vis-à-vis de DFTB plus en détail au sein de la trame principale de cette thèse, mais notons tout de même que GAP semble plus performant que NNP, en particulier aux très courtes distances. Les valeurs GAP de l'énergie de liaison semblent au passage translatées par rapport à VASP; cette surestimation peut-être assez préoccupante, mais elle semble liée à la méthode de référence de GAP (logiciel CASTEP, GGA-PW91, énergie de coupure de la base d'onde planes de 250 eV).

On présente ensuite les densités d'états vibrationnels obtenus pour les monocristaux avec NNP et GAP dans la figure E.2.

Les propriétés vibrationnelles prédites sont médiocres pour la majorité des cristaux; ceci est notamment surprenant pour GAP dans le cas du Silicium β -Sn, car ce cristal est présent dans le jeu d'apprentissage de ce potentiel. On ne note une capacité de prédiction acceptable, pour les deux potentiels, que pour les cristaux diamant et hexagonal. Des deux, le diamant est de loin le mieux représenté; GAP est même parfait pour celui-ci.



FIGURE E.1 – Stabilité relative des différents cristaux de Silicium pur en fonction du volume avec des potentiels purement paramétriques.

NNP est moins précis pour ce système aux modes optiques mais prédit assez bien les modes acoustiques. Les approches purement paramétriques sont donc aptes à impacter sensiblement l'intégralité des fréquences accessibles; cette caractéristique semble hors de portée de notre approche concentrée sur l'énergie répulsive (Chap. 8), uniquement à même d'influer sur la tranche haute des fréquences vibrationnelles ($\omega \ge 7.5$ THz). Enfin, dans le cas du Silicium hexagonal, GAP reproduit très bien les modes optiques au détriment des modes acoustiques et des fréquences intermédiaires, et NNP exhibe exactement le comportement inverse (plus précis pour les modes de acoustiques et sous-estimant la fréquence des modes optiques).

Les résultats précédents sont détaillés et contextualisés plus en détail dans la trame principale du présent document (voir 8.1.3, via la figure 8.4). En particulier, nous y discutons des performances relatives entre les approches purement paramétriques et les potentiels répulsifs DFTB pour la prédictions des propriétés vibrationnelles, et le lien de ces performances avec d'une part les données de références employées pour optimiser les différentes méthodes mais aussi avec les méthodologies elles-mêmes.



FIGURE E.2 – Densités d'états vibrationnels des cristaux de Silicium pur pour des potentiels purement numériques.





FIGURE E.3 – Densité d'états vibrationnels de plusieurs systèmes amorphes de Silicium pour des potentiels purement numériques.

Dans le cas présent, les deux méthodes produisent pour les quatre amorphes étudiés des densités d'états vibrationnels très correctes. On note globalement deux régimes, en fonction du potentiel employé et du domaine de fréquence étudié : pour les modes acoustiques (dominés dans DFTB par les contributions électroniques - Chap. 8), NNP est le plus performant des deux, alors que les modes optiques sont représentées de manière plus similaire par chacun des potentiels, avec peut-être une prédiction légèrement meilleure pour GAP.

Remerciements

Dans ce trop peu de lignes, très maladroites et remplies de superlatifs éculés, je tiens à rendre le crédit qui leur est dû aux nombreuses personnes sans qui ces travaux n'auraient pas abouti. Mon nom figure sur la première page de cette thèse, les leurs méritent tout autant de reconnaissance.

Pour commencer, je tiens à remercier comme il se doit l'école doctorale de Physique et d'Astrophysique et l'Université de Lyon pour m'avoir offert l'opportunité de dédier ces années à la recherche autour de sujets me passionnant. Ma gratitude va également vers l'équipe *Modélisation de la Matière Condensée et des Interfaces* pour leur accueil chaleureux et leur sympathie. Je leur exprime également toute ma reconnaissance pour le mois de contrat supplémentaire qu'ils m'ont accordé : ces quelques semaines m'ont permis d'achever avec rigueur les présents travaux, de parfaire la rédaction de cette thèse et de présenter au terme de ces trois ans un rendu dont la qualité me satisfait. Je tiens enfin à remercier les divers supports administratifs, informatiques, de service et d'entretien dont le travail est indispensable pour que le nôtre soit possible. En particulier, merci à Ruben Mascart grâce à qui les conséquences de la pandémie en cours n'ont pas eu d'impact majeur sur ces travaux.

Je souhaite remercier Thomas A. Niehaus pour son encadrement admirable, sa disponibilité, sa bienveillance et le modèle de rigueur, de savoir et d'ouverture qu'il est pour moi. Diese Thesis könnte immer besser sein, aber brauchte ich genau die Freiheit du mir gegeben hast; vielen dank! Je remercie également Tristan Albaret, dont la curiosité a été une force motrice majeure de ces travaux et de mon évolution personnelle. Il m'est ardu d'exprimer mon admiration tant ta présence et ta compétence m'ont énormément apporté, et même inspiré.

Je tiens aussi à citer Emeric Bourasseau : découvrir la recherche à tes côtés a été une opportunité inestimable. Ton affabilité et ta reconnaissance ont maintenu ma confiance en mes capacités bien au-delà de notre collaboration et jusqu'à ce jour. Je suis très heureux d'avoir pu t'inclure en tant qu'invité au jury de cette thèse, mais il m'est important de t'accorder aussi ces quelques mots supplémentaires. Merci !

Je remercie - *a posteriori* dans ce paragraphe ajouté quelques semaines après la soutenance ! - l'intégralité des rapporteurs et des membres du jury. L'intérêt et l'enthousiasme manifestés lors de leurs divers retours suscitent chez moi tout à la fois une fierté irrépressible (par leur profonde amabilité) et une humilité profonde (par l'éminence de leurs émetteurs et émettrices). Je remercie Olivier Pierre-Louis pour m'avoir donné l'opportunité d'organiser de manière hebdomadaire les réunions d'équipe MMCI : j'ai pu grâce à cela échanger avec des dizaines de personnes brillantes. Je remercie au passage David Martin-Calle pour son soutien capital lors de la préparation de ces séminaires, ainsi qu'Alexandre Nicolas et Kevin Vynck pour la relève qu'ils assurent et le souffle nouveau qui s'en accompagne. Au passage, je remercie amicalement Francesco Boccardo pour les discussions éclairantes, amusantes, et trop peu nombreuses que nous avons eu au fil de cette thèse.

Sur une note plus personnelle, je tiens avant tout à remercier Mme. Jocelyne Bissuel et M. Philippe Bissuel, dont l'amour inconditionnel est jour après jour une source intarissable de motivation. Les chances de me voir arriver ici il y a 25 ans étaient maigres, mais vous m'avez laissé grandir tout en plaçant l'improbable à portée d'effort ; cet accomplissement est le vôtre avant d'être le mien. Il en va de même pour Laurie Bissuel et Corentin Bissuel : vous voir devenir des adultes accomplis, maîtrisant toute difficulté avec cette aisance déconcertante qui vous définit est une inspiration quotidienne. Tâcher d'être digne de l'admiration que vous me portez me fait avancer depuis plus d'une décennie.

A vous ma famille, pour qui le monde universitaire paraît parfois bien abstrait et éloigné de nos préoccupations, merci d'avoir cultivé cette curiosité au point de faire de ma soutenance un évènement important. Votre présence, un jour de semaine, m'apparaissait aussi essentielle qu'insensée : merci !

A vous mes ami.e.s, dont aucun.e ne saisit la portée scientifique de ces travaux; c'est à n'en pas douter la raison vous poussant à vous en enorgueillir infiniment plus que moi. Arriver jusqu'ici m'a forcé à être moins disponible que vous ne le méritez, et je vous dois un Merci majuscule pour votre présence loyale, nécessaire, indéfectible, désintéressée.

Je tiens enfin à figer sur papier mes plus sincères remerciements à Flavie Thérias; nous les relirons tous deux - *sûrement* ! - en souriant dans quelques décennies. Tu m'as soutenu chaque heure de chaque jour au fil des incertitudes, des routes et des appartements qui ont mené à ces lignes, et tu as connu en parallèle de tes propres premières années actives les difficultés quotidiennes d'une thèse autant que n'importe quel jeune docteur. Merci, merci.