

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA
RECHERCHE SCIENTIFIQUE
UNIVERSITE ABOU-BAKR BELKAID DE TLEMCCEN



FACULTE DES SCIENCES

Département de physique

Laboratoire d'Etude et Prédiction de Matériaux
Unité de Recherche de Matériaux et Energies Renouvelables

THÈSE

en vue d'obtenir le grade de Docteur

Présentée

Par

M^r BENYELLOUL Kamel

Thème

**PROJET DU GENOME DES MATERIAUX:
ETUDE DES ALLIAGES D'ACIER PAR LA METHODE
DU DATAMINING ET DE LA DFT**

Soutenue publiquement le 21 Novembre 2013 devant la commission d'examen

- | | | |
|-----------------|--------------------------------------|--------------|
| • G. MERAD | Professeur Université de Tlemcen | Président |
| • H.I. FARAOUN | Professeur Université de Tlemcen | Examinatrice |
| • T. SAHRAOUI | Professeur Université de Blida | Examineur |
| • T. GROSDIDIER | Professeur Université de Metz | Examineur |
| • D. HAMANA | Professeur Université de Constantine | Examineur |
| • H. AOURAG | Professeur Université de Tlemcen | Encadreur |

Remerciements

J'exprime ma plus grande gratitude envers Monsieur Hafid Aourag, mon directeur de thèse, Professeur à l'université de Tlemcen et directeur général de la recherche scientifique et du développement technologique pour m'avoir donné l'opportunité de travailler dans son équipe en me proposant le sujet ainsi que ses conseils avisés tout au long de ces laborieuses et fructueuses années de recherche. Sa patience et ses conseils m'ont permis, quelles que soient les difficultés rencontrées, de mener à bien mon travail de thèse.

Je remercie vivement Mr G. Merad, Professeur à l'université de Tlemcen, d'avoir accepté de présider le jury de thèse.

J'adresse mes sincères remerciements aux membres de jury de thèse, H. Faraoun, Professeur à l'université de Tlemcen, Monsieur T. Sahraoui, Professeur à l'université de Blida, Monsieur D. Hamana Professeur à l'université de Constantine, et Monsieur T. Grosdidier Professeur à l'université de Metz (France), pour tout l'intérêt qu'ils ont porté à mon travail.

J'ai le grand plaisir à remercier mes collègues du Laboratoire d'Etudes et de Prédiction de Matériaux de Tlemcen, ainsi que mes collègues de l'Unité de Recherches Appliquées en Energies Renouvelables de Ghardaia, pour leurs soutien et encouragements.

Je remercie enfin les membres de ma famille, mes parents, qui ont été mes premiers instituteurs, mes frères et sœurs et surtout ma femme pour leur présence à mes cotés dans les moments difficiles de ma vie.

Enfin, je tiens à remercier tous ceux qui, de près ou de loin, ont contribué à la réalisation de ce travail.

TABLE DES MATIERES

Introduction générale.....	1
-----------------------------------	----------

Chapitre I

Etat d'art

1.1	Introduction.....	5
1.2	Généralités sur les aciers inoxydables.....	6
1.3	Classification des aciers inoxydables.....	6
1.3.1	Les aciers inoxydables austénitiques.....	6
1.3.2	Les aciers martensitiques.....	7
1.3.3	Les aciers ferritiques.....	7
1.4	Les éléments d'alliages.....	7
1.5	Approches et modélisations	9
1.5.1	Modélisation par l'approche de la théorie de la DFT.....	10
1.5.2	Modélisation par l'approche du « Datamining ».....	10
1.6	Conclusion.....	12
	Références.....	13

Chapitre II

Théorie de la fonctionnelle de la densité

2.1	Introduction.....	15
2.2	Formalisme de l'équation de Schrödinger.....	15
2.2.1	Equation de Schrödinger.....	15
2.2.2	L'approximation de Born-Oppenheimer.....	17
2.2.3	L'approximation de Hartree.....	18
2.2.4	L'approximation de Hartree-Fock	19
2.3	Principe de la théorie de la fonctionnelle de la densité (DFT).....	20
2.3.1	Les théorèmes de Hohenberg et Kohn.....	20
2.3.1.1	Premier théorème de Hohenberg et Kohn.....	21
2.3.1.2	Deuxième théorème de Hohenberg et Kohn.....	22
2.4	Les équations de Kohn-Sham.....	22
2.5	La fonctionnelle d'échange et corrélation	25
2.5.1	L'approximation de la densité locale (LDA).....	25
2.5.2	L'approximation de gradient généralisé (GGA).....	27
2.6	Méthodes de calcul des structures électroniques	29
2.6.1	Choix de la méthode, de la base des fonctions d'ondes et de la forme du potentiel..	
2.6.2	Théorème de Bloch.....	31
2.6.3	La méthode d'onde planes.....	31
2.6.4	Ondes planes augmentées et ondes planes augmentées linéarisées.....	32
2.6.4.1	Méthode des ondes planes augmentées (APW).....	32
2.6.4.2	Méthode linéaire des ondes planes augmentées linéarisées (LAPW).....	34

2.7	La méthode des pseudopotentiels.....	35
2.8	Le formalisme PAW.....	37
2.9	Conclusion.....	41
	Références.....	42

Chapitre III

Datamining

3.1	Introduction.....	44
3.2	Datamining.....	45
3.3	Techniques et méthodes.....	46
3.3.1	Méthodes non supervisées (techniques descriptives).....	46
3.3.2	Méthodes supervisées (techniques prédictives).....	47
3.4	Réseaux de neurones artificiels	48
3.4.1	Historique.....	48
3.4.2	Neurone formel.....	49
3.4.3	Réseaux de neurones artificiels.....	50
3.4.4	Perceptrons multicouches.....	51
3.4.5	Traitement des données.....	52
3.4.6	Apprentissage des réseaux de neurones.....	52
3.4.7	Rétropropagation du gradient.....	53
3.5	Les algorithmes génétiques.....	58
3.5.1	Principe d'un algorithme génétique.....	59
3.5.2	Représentation des individus et codage.....	60
3.5.2.1	Le codage binaire.....	60
3.5.2.2	Le codage réel.....	60
3.5.2.3	Le codage de Gray.....	61
3.5.2.4	Initialisation	61
3.5.2.5	Fonction d'évaluation.....	61
3.5.2.6	Sharing.....	62
3.5.2.7	Elitisme.....	62
3.5.2.8	Sélection.....	63
3.5.2.9	Croisement.....	64
3.5.2.1	Mutation.....	66
3.6	Avantages et désavantages des l'algorithmes génétiques.....	67
3.7	Conclusion.....	68
	Références	69

Chapitre IV

Application de la DFT

4.1	Introduction.....	72
4.2	Détail de calcul.....	73
4.2.1	Equation d'état.....	73
4.2.2	Module de compressibilité.....	74
4.2.3	Constantes élastiques.....	75

4.3	Conclusion.....	80
	Références.....	81

Chapitre V

Application des techniques du Datamining

5.1	Introduction.....	82
5.2	Application des réseaux de neurones pour la détermination des constantes élastiques de $Fe_{0.62}Cr_{0.185}Ni_{0.185}$	83
5.3	Application d'un système hybride : réseau de neurones et algorithme génétique (AG-RNA).....	92
5.3.1	Application du réseau de neurones.....	93
5.3.2	Application du système hybride (AG-RNA).....	95
5.3.3	Comparaison des résultats.....	97
5.3.3.1	Equation d'état.....	98
5.3.3.2	Module de compressibilité.....	98
5.4	Conclusion.....	99

Conclusion générale

Conclusion et perspectives.....	101
---------------------------------	-----

LISTE DES TABLEAUX

1-1	Différentes nuances des aciers inoxydables austénitiques les plus utilisés.....	7
4-1	Propriétés d'équilibre : constante du réseau a_0 , module de compressibilité B , sa première dérivée B' et le paramètre magnétique $M(\mu_B)$	75
4-2	Constantes élastiques C_{ij} , de l'alliage $\text{Fe}_{0.62}\text{Cr}_{0.185}\text{Ni}_{0.185}$	78
4-3	module de compressibilité (B), module de Young (E), module de cisaillement (G), coefficient de Poisson, la pression de Cauchy ($C_{12} - C_{44}$), et le ratio B/G	79
5-1	Base de données utilisée dans les réseaux de neurones.....	84
5-2	Les constantes élastiques C_{ij} avec l'erreur relative (%).....	90
5-3	Module de compressibilité (B), et le module de cisaillement (G) calculés.....	90
5-4	Base de données utilisée dans le système hybride (AG-RNA).....	93
5-5	Module de compressibilité, calculé par la DFT et AG-RNA.....	99

LISTE DES FIGURES

1-1	Coupe à concentration des alliages ternaire Fe-Cr-Ni :	8
1-2	Diagramme de Schaeffler.....	9
2-1	Organigramme d'un calcul auto-cohérent dans une méthode basée sur la théorie de la fonctionnelle de la densité électronique.....	29
2-2	Les différents traitements de l'énergie cinétique, du potentiel, les fonctions d'ondes.....	30
2-3	Représentation de la cellule unitaire, région interstitielle et la région sphérique....	33
2-4	Représentation schématique du remplacement d'une fonction d'onde φ et du potentiel $V(r)$ par une pseudo fonction d'onde φ_{pseudo} et un pseudo potentiel V_{pseudo} respectivement.....	35
3-1	Schéma du processus du data mining.....	46
3-2	Exemple de techniques non supervisées (Analyse en Composantes Principale).....	47
3-3	Exemple de Techniques supervisées (Arbre de décision)	47
3-4	Un neurone formel.....	49
3-5	Différentes fonctions d'activation usuelles.....	50
3-6	Réseau de neurones non bouclé.....	50
3-7	Réseau de neurones bouclé.....	51
3-8	Schéma synoptique du procédé d'apprentissage du réseau de neurone.....	52
3-9	Modèle du neurone j.....	53
3-10	Principe général des algorithmes génétiques standard.....	59
3-11	Exemple de codage réel.....	60
3-12	Schéma de la roulette de sélection	63
3-13	Croisement à un point.....	65
3-14	Croisement à deux points.....	65
3-15	Croisement uniforme.....	65
3-16	Opérateur de mutation.....	66
4-1	Structure cristalline du Fer.....	72
4-2	Structure cristalline du $Fe_{10}Cr_3Ni_3$	73
4-3	Variation de l'énergie totale en fonction du volume de Fe et de $Fe_{0.62}Cr_{0.185}Ni_{0.185}$...	74
5-1	Structures de réseau de neurones à trois sorties.....	85
5-2	Entraînement de convergence du réseau de neurones pour C_{11}, C_{12} et C_{44}	86
5-3	Structures de réseau de neurones à une seule sortie (C_{11} ; C_{12} ou C_{44}).....	86
5-4	Entraînement de convergence du réseau de neurones pour C_{11}	87
5-5	Entraînement de convergence du réseau de neurones pour C_{12}	87
5-6	Entraînement de convergence du réseau de neurones pour C_{44}	88
5-7	Résultats du modèle de régression du réseau de neurones comparés avec les valeurs de la base de données pour les constantes C_{11} et C_{12}	89

5-8	Résultats du modèle de régression du réseau de neurones comparés avec les valeurs de la base de données pour la constante C_{44}	89
5-9	Structures de réseau de neurones à deux sorties (C_{11} et C_{12}).....	91
5-10	Entraînement de convergence du réseau de neurones pour C_{11} et C_{12}	91
5-11	Structures du réseau de neurones utilisées (3-couche cachée-1).....	94
5-12	Entraînement de convergence du réseau de neurones pour B	95
5-13	Résultats du modèle de régression du réseau de neurones pour B	95
5-14	Organigramme d'un modèle hybride (AG-RNA).....	96
5-15	Evolution de la fonction fitness en fonction des générations.....	97
5-16	Structure cristalline du $\text{Fe}_8\text{Cr}_3\text{Ni}_3\text{Mn}_2$	98
5-17	Variation de l'énergie totale en fonction du volume de $\text{Fe}_{0,50}\text{Cr}_{0,185}\text{Ni}_{0,185}\text{Mn}_{0,125}$	98

INTRODUCTION GENERALE

INTRODUCTION GENERALE

L'amélioration des matériaux existants et le développement de nouveaux matériaux sont reconnus comme étant un des domaines-clés de l'essor industriel. Le développement de la science des matériaux a conduit les chercheurs à comprendre que les propriétés d'un matériau dépendaient étroitement de sa structure et de son environnement. Leurs investigations scientifiques sont généralement menées suivant les deux axes classiques (i) la simulation et (ii) l'expérimental. Un certain nombre d'outils de simulation sont élaborés, permet d'étudier des systèmes physiques de différentes natures (atomes, molécules ou solides). Parmi elles, la théorie quantique de la fonctionnelle de la densité (*Density functionnelle theory* ; DFT) qui est une reformulation de la mécanique quantique, où il est montré que la densité électronique de l'état fondamental est une grandeur tout aussi fondamentale que la fonction d'onde et dont la détermination permet de calculer toutes les propriétés du matériau étudié. La DFT a été popularisée par Walter Kohn et L. J. Sham en 1965 grâce au formalisme de Kohn-Sham, qui utilise un jeu d'équations comparables à celles d'Hartree-Fock, et surtout l'introduction des fonctionnelles d'échange-corrélation modélisant la corrélation électronique. Différentes fonctionnelles existent pour la DFT, telle que l'approximation de la densité locale (LDA) et le gradient généralisé (GGA). De nos jours, la DFT est la méthode de calcul de propriétés physiques et chimiques la plus utilisée, car elle permet de traiter la corrélation des systèmes comprenant un nombre important d'électrons. Mais cette méthode présente une certaine limitation, car cette théorie ne permet d'étudier que des matériaux idéalisés, et ignore les complexités des formulations qui sont généralement présents dans les matériaux réels, et trouve des difficultés relatives à les modéliser.

La recherche de nouveaux matériaux aux propriétés physiques spécifiques telles que, une dureté importante, une bonne ductilité et une grande résistance à la corrosion pour un fonctionnement dans des conditions extrêmes, revient à utiliser l'une ou les deux classes de recherche citées (théorique et expérimentale) afin de développer des matériaux *exotiques*. Cependant, l'application de l'une ou l'autre peut ne pas toujours être concluante et le matériau synthétisé peut ne pas répondre aux exigences souhaitées.

Un nouvel axe sur la conception des matériaux peut être inclus dans le cadre de cette recherche. Ce travail s'inscrit dans le cadre d'un projet de recherche portant sur le « *génom*e des matériaux », créé au sein du Laboratoire d'Etude et Prédiction de Matériaux de

l'Université de Tlemcen dans le but de donner une nouvelle façon de voir, d'enseigner la science et d'accentuer le développement technologique dans l'espoir de révolutionner les connaissances dans le domaine de la science des matériaux.

Dans les sciences génétiques, chaque individu vivant d'une même espèce est porteur de gènes (la corpulence, la couleur de la peau, la couleur des yeux, la beauté, la force, ...etc.), ces gènes peuvent être des qualités ou des défauts. L'idée du projet est de transférer le concept du vivant à l'inerte (la matière inerte) en utilisant quelques outils et mécanismes de l'évolution des espèces, telles que la mutation, la sélection et la reproduction. Etant donné que la technologie n'est pas assez évoluée, et que les moyens à mettre en évidence d'une telle recherche sont importants, nous sommes dans une situation où il faudra accélérer l'idée sans passer ni par l'expérience ni par des calculs théoriques, il s'agit d'accélérer la découverte en se basant juste sur le savoir acquis jusqu'à présent. C'est ce qu'on appelle l'extraction de connaissances à partir de données où « Datamining ».

Par ailleurs, depuis l'explosion des capacités de stockage informatique au moins à partir du début des années 1990, la question de l'analyse de grands volumes de données s'est imposée. Basée sur l'extraction de connaissances et l'analyse de données, permet d'obtenir des informations synthétiques à partir de données recueillies. L'extraction de connaissance à partir de données (ECD) ou en anglais, *Knowledge Discovery in databases (KDD)* est composée de quatre phases : (i) acquisition et stockage des données (Data Warehousing), (ii) pré-traitement des données (Pré-processing), (iii) fouille de données (Data Mining) et (iv) post-traitement (Post-Processing). La troisième phase est le domaine de recherche qui tente de répondre à la question de l'analyse d'un grand volume de données en mettant au point des outils de fouille de données. Nos travaux se situent principalement au niveau de cette étape, elle intègre à la fois le choix de la modélisation adéquate, la méthode à utiliser ainsi que son application à la recherche de structures de données sous-jacentes et à la création de modèles explicatifs et prédictifs.

Fouiller des données ne sert à rien si les résultats de cette fouille ne sont pas interprétables. La fouille de données n'est donc qu'une étape d'un processus d'extraction de connaissances plus large qui s'étend de la préparation des données jusqu'à l'interprétation des résultats. La fouille de données établie par ce biais des connections vers d'autres disciplines préexistantes telles que les bases de données, les outils de visualisation, les modèles de représentation de la connaissance et plus largement l'intelligence artificielle. Résoudre un problème par le processus d'exploration de données impose généralement l'utilisation d'un grand nombre de méthodes et d'algorithmes différents. Les méthodes de fouilles de données

sont nombreuses et varient principalement en fonction d'une part, de la nature de la quantité des données considérées et d'autre part, des questions auxquelles on cherche à répondre.

L'objectif de notre thèse est l'application de différents outils d'extraction de connaissances à partir de données dans le domaine de la science des matériaux, afin d'aboutir à une compréhension plus ample des propriétés physiques des matériaux et alliages. Les résultats trouvés par les techniques du Datamining seront comparés à ceux déterminés dans un premier temps par la méthode de simulation basée sur la DFT. Notre choix s'est porté essentiellement à l'étude et la détermination de certaines propriétés physiques (constantes élastiques et module de compressibilité) des alliages des aciers inoxydables, à savoir l'alliage ternaire $Fe_{0.62}Cr_{0.185}Ni_{0.185}$ et le quaternaire $Fe_{0.50}Cr_{0.185}Ni_{0.185}(Mn)$.

Cette thèse est organisée comme suit :

Dans le premier chapitre, un état d'art sur les aciers, ses différentes applications, ainsi que ses différentes nuances les plus utilisées selon la norme AISI (American Iron and Steel Institute) seront présentés. Dans l'étape suivante, nous présentons un aperçu sur deux méthodes d'approches et de modélisations, la première basée sur la théorie de la fonctionnelle de la densité « DFT » qui peut apporter des informations cruciales et aider à comprendre des lois de comportement d'origine microscopique afin de guider plus efficacement le choix des matériaux pour une propriété souhaitée. La seconde, basée sur l'analyse et l'exploration des données, dénommée aussi « Datamining ».

Le deuxième chapitre sera consacré à l'étalage des principes fondamentaux de la théorie de la fonctionnelle de la densité (DFT) en soulignant les différentes approximations. Nous avons donné aussi un aperçu sur la fonctionnelle d'échange et de corrélation. Dans l'étape suivante, nous présentons les deux approches permettant l'application de cette théorie, à savoir l'approximation d'ondes planes, et la deuxième approche, l'approximation du pseudopotentiel pour laquelle nous décrivons le principe.

Le troisième chapitre est consacré à la définition du datamining et ses différentes techniques. En premier temps, nous allons donner une représentation générale sur les réseaux de neurones, nous présentons ensuite les deux grandes familles de structures neuronales les plus utilisées (perceptron multicouche), nous abordons également le problème d'apprentissage des paramètres de ces structures ainsi que le traitement des données. Dans un second temps, nous allons définir les algorithmes génétiques, leurs principaux fonctionnements et nous faisons une description détaillée de ses différentes composantes.

En chapitre 4, nous abordons les calculs théoriques par la méthode de la théorie de la fonctionnelle de la densité (DFT) implémentée dans le code de calcul Vienna Ab-initio Simulation Package (VASP). Nous avons choisi d'étudier dans un premier temps les propriétés structurales du Fer et de l'alliage inoxydable $Fe_{0.62}Cr_{0.185}Ni_{0.185}$. Dans un second temps, nous allons déterminer les propriétés élastiques de l'alliage inoxydable $Fe_{0.62}Cr_{0.185}Ni_{0.185}$ par la méthode de la conservation du volume de la maille sous l'effet de déformations (*volume conserving distortions methods*). Au cours de cette étude, nous allons confronter ces prédictions aux résultats déjà acquis expérimentalement ainsi qu'aux travaux théoriques consacrés à cet égard.

Le cinquième chapitre offre une application des approches fondées sur le processus d'extraction de connaissances à partir des données, appelé également « Datamining ». Pour ce faire, une base de données a été conçue et sera exploitée par différentes techniques du Datamining. Dans un premier temps, nous présentons une application des réseaux de neurones afin de modéliser la relation composition - propriétés élastiques en utilisant plusieurs structures de réseau de neurones du type perceptron multicouche. Par la suite, les trois constantes élastiques C_{ij} de l'acier inoxydable ternaire $Fe_{0.62}Cr_{0.185}Ni_{0.185}$ seront déterminées et comparées à des résultats expérimentaux et théoriques basés sur la DFT. Dans un second temps, une nouvelle approche formée d'une hybridation entre le réseau de neurones et l'algorithme génétique (GA-RNA) sera présentée et appliquée. Cette structure hybride nous permettra d'évaluer la concentration de la composition chimique optimale de l'élément manganèse dans l'alliage quaternaire $Fe_{(bal)}Cr_{0.185}Ni_{0.185}(Mn)$ et de prédire le module de compressibilité (B) correspondant.

Enfin, nous terminons notre travail par une conclusion générale, qui illustre les principaux résultats obtenus en tirant quelques perspectives que nos approches offrent dans le domaine de recherche sur les alliages métalliques, et qui ouvrent de nombreuses voies à la résolution des problèmes qui peuvent être envisagés.

CHAPITRE I

ETAT D'ART

ETAT D'ART

1.1 Introduction

La recherche de nouveaux matériaux aux propriétés physiques et/ou chimiques spécifiques est un enjeu majeur de l'industrie actuelle, et ce quels que soient les domaines d'applications considérés (énergie, constructions naval et aéronautiques,... etc.).

Un des meilleurs matériaux que l'on puisse donner comme exemple est l'acier. Ce matériau attire une attention croissante des chercheurs en raison de ses applications multiples. L'acier est un alliage composé essentiellement de Fer (Fe) très majoritaire et de carbone (C) dont les proportions sont comprises entre 0.02% et 2% en masse. C'est essentiellement la teneur en carbone qui confère à l'alliage les propriétés du métal qu'on appelle « *acier* ». Car il existe d'autres alliages à base de Fer qui ne sont pas considérés comme tel, la fonte et les ferronickels par exemple.

L'intérêt majeur des aciers réside dans leurs propriétés mécaniques satisfaisantes tel qu'une forte résistance à la déformation élastique, résistance au choc, néanmoins on peut leur reconnaître quelques inconvénients, notamment leur mauvaise résistance à la corrosion, à laquelle on peut toutefois remédier, soit par divers traitements de surface (peinture, galvanisation à chaud, brunissage) soit par ajout d'autres éléments d'alliages pour obtenir un acier inoxydable. Les aciers inoxydables, couramment appelés « *inox* » sont des aciers auxquels on ajoute essentiellement du chrome, qui au delà de 10.5 % en solution, provoque une formation d'une couche protectrice d'oxyde de chrome qui leur confère l'inoxidabilité [1-3]. Car l'une des principales propriétés de l'acier inoxydable est sa résistance remarquable à la corrosion. D'autres éléments peuvent être ajoutés, notamment le nickel, qui améliore les propriétés mécaniques en général et la ductilité en particulier.

Les premiers aciers résistants à base de chrome furent développés par *Pierre Berthier* qui remarqua leur résistance à certains acides, mais la présence de carbone en grande quantité rendait le matériau fragile. En 1890, *Hans Goldschmidt* développa et breveta un procédé appelé « *la thermit* » qui permettait d'obtenir du fer sans carbone. Entre 1904 et 1911, divers chercheurs, notamment *Léon Guillet*, ont mis aux point différents alliages d'aciers que l'on pourrait aujourd'hui considérer comme inoxydables. En 1911, *Philip Monnartz* mettait en évidence l'influence du chrome sur les alliages d'aciers et leur résistance à la corrosion. Et ce n'est qu'en 1913 que l'anglais *Harry Brearley* des laboratoires *Brown-Firth*, en travaillant sur

des pièces de canons d'armes à feu, développa un acier qu'il baptisa d'abord *rustless* « sans rouille », il s'aperçut que des échantillons polis en vue d'examen de laboratoire ne subissaient pas d'oxydation. Cet acier sera rebaptisé *stainless* 'sans taches'. Et ce fut la première fois qu'on parle d'acier inoxydable. Depuis les années 70, de nouveaux aciers inoxydables commencèrent à être développés. Ils ont une meilleure résistance à la corrosion. C'est à ce moment que l'utilisation de ce matériau commença à augmenter.

1.2 Généralités sur les aciers inoxydables

Les aciers inoxydables jouent un grand rôle dans d'innombrables domaines : vie quotidienne, industrie mécanique, agroalimentaires, chimie, transport...etc. [4-6]. La construction est le premier secteur de son utilisation, il permet de réaliser des œuvres extrêmement hardis qui peuvent supporter des charges les plus lourdes et couvrir avec des sections réduites des portées immenses soit en longueurs (constructions de ponts ; pont de Golden Gates d'une portée de 1 200 m), soit en hauteur (construction des gratte-ciel : Empire State building d'une hauteur de 379 m). Utilisé aussi en médecine pour sa propriété biocompatible, c'est-à-dire des matériaux que les cellules vivantes acceptent et intègrent dans la structure tissulaire.

1.3 Classification des aciers inoxydables

Il existe différents types d'aciers inoxydables suivant leur teneur en chrome et les traitements thermiques qu'ils ont subis, ils présentent une large gamme de propriétés. Dans la suite nous allons brièvement décrire les différents types d'aciers inoxydables [7].

1.3.1 Les aciers inoxydables austénitiques

Les aciers inoxydables austénitiques contiennent généralement entre 0.02 % et 0.1 % de carbone, 17 à 20% de chrome, 7 à 25 % de nickel et 2 à 5 % de molybdène et d'autres éléments d'addition comme le cuivre, le silicium, le titane et le niobium [8,9].

La teneur en nickel sert à donner et à conserver la structure austénitique à ces aciers même à températures ambiante, en plus d'augmenter leur résistance à la corrosion. Les aciers austénitiques sont ductiles, résistants au choc, ils ne se trempent pas, mais peuvent acquérir une certaine dureté par déformation à froid. A l'état recuit, ils possèdent une plus grande ductilité et une meilleure ténacité que les aciers ordinaires. Ils possèdent une très bonne

résistance à la corrosion. Le tableau 1-1 liste différents aciers inoxydables austénitiques les plus connus et les plus employés [10].

Tableau 1-1 : Différentes nuances des aciers inoxydables austénitiques les plus utilisés.

AISI	C	Mn	P	S	Si	Cr	Ni	Mo	Fe
304	0.08	2	0.045	0.03	1.0	18-20	8-10	-	Balance
310	0.25	2	0.045	0.03	1.5	24-26	19-22	-	Balance
316	0.08	2	0.045	0.03	1.0	16-18	10-14	2-3	Balance
317	0.08	2	0.045	0.03	1.0	18-20	11-15	3-4	Balance

AISI: American Iron and Steel Institute.

1.3.2 Aciers martensitiques

Ces aciers contiennent en général 12 à 19% de chrome, leur teneur en carbone varie entre 0.08 et 1.2% [8] ; ils peuvent contenir du nickel et du molybdène ainsi que certains éléments d'addition tels que le cuivre, le titane ou le vanadium. Dans la pratique on les utilise après un traitement thermique de trempe et revenu assurant ainsi un compromis entre résistance -résilience - tenue à la corrosion.

1.3.3 Aciers ferritiques

Leur teneur en chrome est de 12 à 20 %. Les aciers inoxydables ferritiques sont moins résistants et moins ductiles que les aciers austénitiques, leur résistance mécanique peut être augmentée par écrouissage [8].

1.4 Les éléments d'alliages

Les éléments d'alliages et leurs proportions permettent d'améliorer la tenue à la corrosion et les propriétés mécaniques du matériau. Ils sont classés en deux grandes familles [6] :

La première famille des éléments alphagène : ces éléments favorisent la prépondérance à la température ambiante d'une structure cristallographique cubique centrée. Parmi ces éléments le chrome, sont accroissement favorise la formation et la stabilité de la ferrite δ au détriment de l'austénite γ (figure1-1), il y a notamment le molybdène, le titane, le silicium, et le niobium qui sont aussi des éléments alphagène.

La seconde famille est celle des éléments gammagène : ils favorisent la prépondérance à la température ambiante d'une structure cristallographique cubique à face centrée. Parmi ces

éléments le nickel qui élargit le domaine de l'austénite γ (figure 1-1) au détriment de la ferrite et facilite la mise en œuvre des aciers inoxydables par déformation plastique. Il ya notamment le carbone, l'azote, le manganèse qui appartiennent à la famille des gammagène [11].

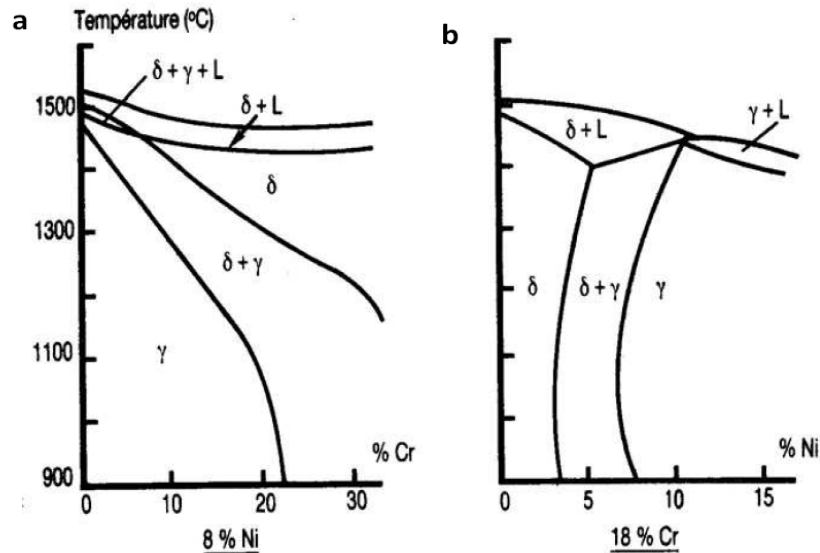


Figure 1-1 : Coupe à concentration des alliages ternaires Fe-Cr-Ni :

- (a) mise en évidence des caractères alphas du chrome,
- (b) mise en évidence des caractères gammagènes du nickel

Partant du constat que certains éléments d'alliage ont un effet alphas, c'est-à-dire qu'ils favorisent la formation d'une structure cristallographique cubique centrée comme le chrome et que d'autres ont un effet gammagène, c'est-à-dire qu'ils favorisent la formation d'une structure cristallographique cubique à faces centrées comme le nickel. Il existe plusieurs diagrammes qui ont été proposés pour déterminer la structure de ces alliages métalliques. Le plus connu est celui de Schaeffler [12], celui ci permet une représentation des différentes structures des aciers (Fe-Cr-Ni) et de les prévoir (figure1-2) en attribuant un coefficient d'équivalence avec le chrome ou le nickel à chaque élément d'addition, en fonction de son pouvoir alphas, ou gammagène. Traditionnellement, les effets de différents éléments d'alliage ont été quantifiés par l'utilisation du nombre de chrome et nickel équivalent. Plusieurs formules ont été proposées par Schaeffler [12], Delong [13] et Hull [14]. L'expression donnée par Rhodes et al [15].

$$Ni_{eq} = \% Ni + 0.65\% Cr + 0.98\% Mo + 1.05\% Mn + 0.35\% Si + 12.6\% C. \quad (1.1)$$

$$Cr_{eq} = \% Cr + \% Mo + 1.5(\% Si) + 0.5 (\% Nb) \quad (1-2)$$

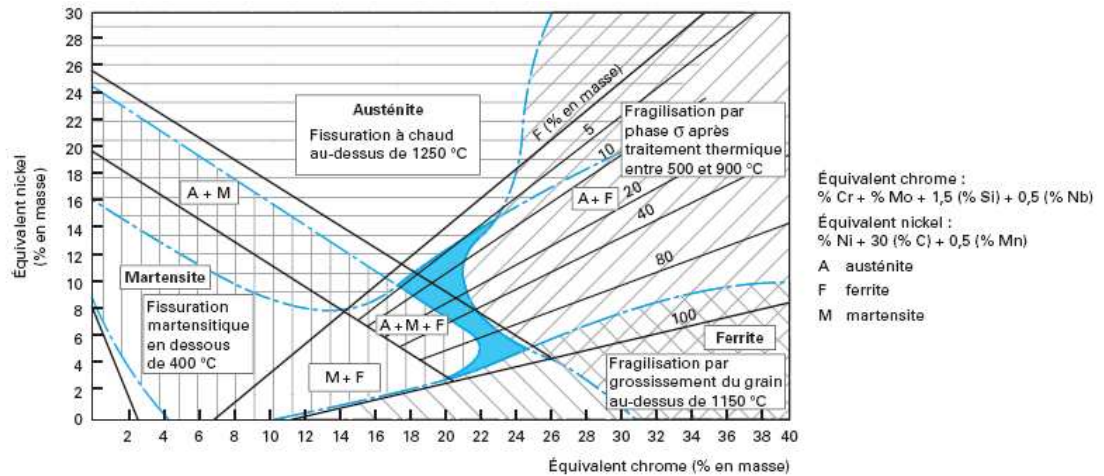


Figure 1-2 : diagramme de Schaeffler.

Le développement des aciers inoxydables austénitiques est lié à la recherche des nuances des aciers, que ce soit par des méthodes théoriques ou par des procédés expérimentaux. Ces méthodes permettent d'avoir une vision détaillée sur les interactions entre les atomes et d'augmenter ainsi leurs potentialités dans le but d'obtenir des alliages performants en termes de propriétés mécaniques et de résistance à la corrosion, et une large application technologiques.

1.5 Approches et modélisations

Une compréhension fondamentale de la relation entre les éléments d'alliage et les propriétés mécaniques nécessite une approche synergique impliquant à la fois la théorie et l'expérience. La théorie peut être utilisée non seulement pour comprendre les résultats expérimentaux, mais aussi à orienter les futures expériences. Les grandes avancées dans la méthodologie (théories et algorithmes) et l'augmentation des puissances de calcul au cours de ces dernières années ont ouvert de nouvelles possibilités pour les études théoriques des matériaux en général et les alliages d'aciers en particulier. Une approche théorique très élaborée est nécessaire pour compléter les efforts expérimentaux dans la recherche fondamentale sur la conception et la synthèse de nouveaux matériaux. Le défi pour les approches théoriques est la modélisation. D'une manière générale deux catégories d'approches théoriques sont utilisées pour l'étude des systèmes physiques :

- L'approche de la mécanique quantique, qui donne des informations sur la structure électronique et sur les liaisons chimiques
- L'approche issue d'exploitation de données « *Datamining* »

1.5.1 Modélisation par l'approche de la théorie DFT

Elle s'avère indispensable pour une meilleure compréhension des phénomènes physiques. Cette méthode a pour but de guider les expérimentateurs dans leurs recherches de nouveaux matériaux ou plus performants. Sur le plan théorique il s'agit d'apporter un éclairage au niveau fondamental sur la relation des microstructures et leurs propriétés au moyen d'étude *ab-initio* dans le cadre de la théorie de la fonctionnelle de la densité (DFT). Par son application pour les matériaux en général et les aciers inoxydables en particulier, cette méthode peut apporter des informations cruciales et aider à comprendre des lois de comportement d'origine microscopique, afin de guider plus efficacement le choix des matériaux pour une propriété souhaitée.

Sur des travaux basés sur la DFT, *Levente Vitos et al* [16] ont récemment étudié le comportement et l'évolution du module de compressibilité (B) et le module de cisaillement (G) en fonction de la teneur du chrome (Cr) et du nickel (Ni) des alliages aciers inoxydables ternaire $Fe_{100-(n+c)}Cr_cNi_n$ avec $135 \leq c \leq 25.5$ et $8 \leq n \leq 24$. Ces derniers ont montré que ces modules d'élasticité sont en fonction de la microstructure et ont mis l'accent sur le comportement des propriétés élastiques en fonction des différentes concentrations de l'élément Cr et l'élément Ni. Parmi d'autre travaux basés sur l'approche de la mécanique quantique on peut citer ceux de *V.Rajevac et al* [17] pour la détermination des différents modules d'élasticité sur les aciers Fe-18Cr-12Ni-2Mo, et Fe-18Cr-16Ni-10Mn. On peut citer aussi les travaux de *B. Johansson et al* [9] sur l'utilisation de la DFT pour la détermination des constantes élastiques de différents aciers.

1.5.2 Modélisation par l'approche du « Datamining »

Dans le domaine de la modélisation, le développement de nouvelles méthodes qui permettent la compréhension des facteurs qui contrôlent les mécanismes des propriétés mécaniques est primordial actuellement. Les moyens informatiques modernes ont permis de produire et d'archiver d'énormes masses de données numériques depuis maintenant au moins deux décennies. Si ces données sont généralement collectées pour rendre un service donné ou répondre à une question précise, éventuellement ces données renferment aussi de nombreux éléments de connaissance relatifs aux objets qui sont décrits. Le problème de l'accès à cette connaissance dépasse toutefois largement les capacités humaines d'analyse tant ces éléments de connaissances sont disséminés dans une quantité importante de données souvent complexe. La mise à disposition de grandes quantités de données d'une part et l'impossibilité de les exploiter pleinement d'autre part, ont favorisé dès le début des années 90 l'essor d'une

nouvelle discipline scientifique appelée tantôt extraction de connaissance à partir de données [18,19] par la communauté d'intelligence artificielle, tantôt fouille de données [20,21] par la communauté des bases de données.

Cependant, grâce au développement et aux efforts réalisés ces dernières années et l'élaboration de nouveaux algorithmes tels que les réseaux de neurones [22] et les algorithmes génétiques [23] permettant aux chercheurs d'exploiter et de découvrir des modèles de plus en plus précis. Cette approche doit être en mesure de contribuer de manière utile et significative au développement de nouveaux matériaux et l'amélioration des matériaux existants.

L'idée maîtresse de cette approche est de chercher à extraire un savoir ou d'une connaissance à partir de données par des méthodes basées sur des algorithmes. Ces données portant un ensemble d'informations sur différents matériaux qu'on veut étudier (composition chimiques, constantes élastiques, propriétés mécaniques), issues à partir des études fondamentales (théoriques et expérimentales). L'analyse des données permet d'obtenir des informations synthétiques à partir de données recueillies.

Le défi du Datamining, consiste à utiliser différentes techniques afin de concevoir un modèle d'un phénomène [24] (propriétés mécaniques) complexe, pour mieux le comprendre, afin de réduire les coûts de recherches ou d'exploitations liés à ce phénomène, ou bien de pouvoir l'expliquer de façon plus efficace et plus rigoureuse. Comme par exemple établir un modèle qui puisse prédire un comportement d'une propriété physique en fonction des compositions chimiques.

Cette idée qui s'est concrétisée à travers plusieurs travaux dans le domaine de la science des matériaux par plusieurs chercheurs, ouvre une perspective intéressante sur l'utilisation des algorithmes qui rentre dans le cadre de l'extraction de connaissances à partir des bases de données. *Bahrami et al* [25,26] ont utilisé les réseaux de neurones afin de trouver des modèles de corrélations entre les composition chimiques et la limite d'élasticité, la résistance à la traction. Aussi, l'utilisation des réseaux de neurones a permis à *M.Boltani Esfahani et al* [27] de modéliser d'autres propriétés mécaniques en fonction des différentes composition chimique des alliages d'aciers à faible teneur de carbone. On peut citer a titre d'exemple et dans le cadre de la prédiction du nombre de ferrite par les réseaux neuroniques, les travaux de *M.Vasudeven et al* [28].

1.6 Conclusion

Dans ce chapitre, nous avons présenté un état d'art sur les aciers en général et les aciers inoxydables en particulier, leurs classifications ainsi que leurs applications. Nous avons aussi présenté des notions sur différentes approches de modélisation et leurs applications dans le domaine de la science des matériaux, à savoir la théorie de la fonctionnelle de la densité (DFT) et celle basée sur l'exploitation et l'extraction de connaissances à partir des bases de données « Datamining » qui seront exposées avec plus de détail pour chacune dans les chapitres suivants.

Références

- [1] A.Shahryari, S.Omanovic,J.A.Szpunar, Materials Science and Engineering C28 (2008) 94.
- [2] J. Okado, K. Okada, A. Ishiyama, Y. Setsuhara, K.Takenaka, Surface and Coatings Technology 202 (2008) 5595.
- [3] C.C.Shih, C.M. Shih, Y.Y. Su, L.H.J.Su, M.S.Chang, S.J.Lin, Corrosion Science 46 (2004) 427.
- [4] P. Lours, G. Bernhart, T. Cutard, D. Delagnes, V. Velay « Comportement, Durabilité et structure des matériaux (CDSM) » 2006.
- [5] I. Fonteneau, J. Ledion, G. Nouail, « aciers inoxydables, corrosion- protection des réseaux » ASTEE 2006.
- [6] Thèse Mahmoudi S, « Etude comparative entre deux métaux passifs en milieux neutre alcalin et acide : un acier inoxydable 304L et un alliage de Titanr Ti6Al4V » université de Constantine (2008).
- [7] K.H. Lo, C.H. Shek, J.K.L. lai, Materials Science and Engineering R R65 (2009) 39-104.
- [8] Aciers inoxydables, Critères de choix de structure, Technique et l'ingénieur, traité Matériaux métallique Vol MB4, M4540, Paris.
- [9] S. Krakowiak, K. Darowicki, P. Slepski, Electrochimica Acta 50 (2005) 2699-2704.
- [10] B.Johansson, L.Vitos, P.A.Korzhavyi, Solid State Science 5 (2003) 931-936.
- [11] Bavay J.C. « Aciers inoxydables austénitiques », les aciers inoxydables, Editions de physique, p.565-609,1990.
- [12] A.L. Shaeffler, Metal Progress 56 (1949)680.
- [13] C.J. Long, W.T. DeLong, Welding Journal 52 (1973) 281s.
- [14] F.C. Hull, Welding Journal 52 (1973) 183s.
- [15] C.G. Rodes, A.W. Thompson, Metallurgical Transactions A8 (1977) 1901.
- [16] L.Vitos, P.A. Korzhavyi, Doi:10.1103/PhysRev Lett,88.155501.
- [17] V. Rajevac , M. Hoelzel, S.A. Danilkin, A. Hoser, H. Fuess, Journal of Physics Condensed Matter 16 (2004) 2609-2616.
- [18] D.J. Lubinsky, Discovery from databases : A review of AI and statistical techniques. In proceedings of IJCAI-89 workshop on knowledge-Discovery in Databases, p. 204-218, 1989.

-
-
- [19] G.Piatetsky-Shapiro et W. Frawley, édés knowledge-Discovery in Databases AAAI Press/MIT Press, Menlo Park (Ca) and Cambridge (Ma), 1991.
- [20] T.Anwar, H.Beck, S.Navathe : Knowledge mining by imprecise querying : A classification-based approach. IEEE intl. Conf. on data Eng, p 622-630,1992.
- [21] R.S.Michalski, L.Kerschberg, K.A.Kaufmann et J.S.Ribeiro: Mining for knowledge in databases: the Inlen architecture, initial implementation and first results. J.Intell.Inf Syst, 1(1)85-133, 1992.
- [22] Mark W.Craven, Jude W. Shavlik, Future Generation Computer Systems 13 (1997)211-229.
- [23] K.G.Srinivasa, K.R.Venugopal, L.M.Potnaik, Information Sciences 177 (2007) 4295-4313.
- [24] U. Fayad et al. Advances in Knowledge Discovery and Datamining (AAAI/MIT-press 1995)
- [25] A. Bahrami, S.H.Mousavi Anijdan, A.Ekrami, Journal of Alloys and Compounds 392 (2005)177-182.
- [26] M.Yazdanmehr, S.H.Mousavi Anijdan, A.Samadi, A. Bahrami, Computational Materials science 44 (2009) 1231-1235.
- [27] M. Boltani Esfahani, M.R. Toroghinejad, A.R. Yeganeh, Mat. and Des. 30 (2009) 3653-3658.
- [28] M. Vasudeven, B.P.C. Ras, B. Venkatraman, T.J. Jayakumer, Baldev Raj, Journal of Materials Processing Technology 169 (2005) 396-400.

CHAPITRE II

THEORIE DE LA FONCTIONNELLE

DE LA DENSITE

THEORIE DE LA FONCTIONNELLE DE LA DENSITE

2.1 Introduction

La théorie de la fonctionnelle de la densité DFT pour « *Density Functional Theory* », fut introduite au milieu des années soixante par *Hohenberg* et *Kohn* [1], *Kohn* et *Sham* [2]. La contribution de *Walter Kohn* par la théorie de la fonctionnelle de la densité à la compréhension des propriétés électroniques, en particulier en physique de la matière condensée, a été récompensée par le prix Nobel de chimie en 1998. Cette théorie permet en effet une application efficace des principes de base de la mécanique quantique dans le code de calcul numérique dit *ab initio* pour déterminer les propriétés électroniques d'un groupement atomiques. En fait la connaissance des propriétés électroniques permet d'obtenir des informations sur les caractéristiques structurales (mécaniques, élastiques ...etc.).

Le but de ce chapitre est d'introduire la notion de la DFT dans le calcul de l'énergie de l'état fondamental et les propriétés élastiques de certains alliages objet de notre travail.

2.2 Formalisme de l'équation de Schrödinger

2.2.1 Equation de Schrödinger

Les solides sont constitués par une association de particules élémentaires, les ions et les électrons. Le problème théorique fondamental de la physique des solides est de comprendre l'organisation intime de ces particules à l'origine de leurs propriétés. Mais dans ce cas la mécanique classique s'avère être insuffisante et il faut faire appel à la mécanique quantique dont la base est la résolution de l'équation de Schrödinger [3] qui s'écrit sous la forme connue :

$$H\psi = E\psi \quad (2-1)$$

Où H est l'hamiltonien total du système contenant les termes d'énergie cinétique et potentielle apportées par les électrons et les noyaux dont leurs nombres est respectivement N_e et N_n , E étant l'énergie totale du cristal et ψ la fonction d'onde totale du système qui est fonction des coordonnées r_i des électrons et R_n des noyaux. Cette fonction d'onde contient toutes les informations du système.

$$\psi = \psi(r_1, r_2, \dots, R_1, R_2, \dots) \quad (2-2)$$

L'opérateur hamiltonien exact du système (non relativiste) peut s'écrire comme la somme de toutes les formes d'énergies :

$$H = T_n + V_{nn} + V_{ne} + V_{ee} + T_e \quad (2-3)$$

Où l'on définit les termes comme suit :

- Energie cinétique des électrons :

$$T_e = \sum_{i=1}^{N_e} T_i = \sum_{i=1}^{N_e} \frac{-\hbar^2}{2m} \Delta_i \quad \text{avec } m \text{ la masse électronique} \quad (2-4)$$

- Energie cinétique des noyaux :

$$T_n = \sum_{\lambda=1}^{N_n} T_\lambda = \sum_{\lambda=1}^{N_n} \frac{-\hbar^2}{2M} \Delta_\lambda \quad \text{avec } M \text{ la masse d'un noyau} \quad (2-5)$$

- Energie d'interaction électrons-électrons :

$$V_{ee} = \frac{1}{2} \sum_{\substack{i,j=1 \\ i \neq j}}^{N_e} \frac{e^2}{|r_i - r_j|} = \frac{1}{2} \sum_{\substack{i,j=1 \\ i \neq j}}^{N_e} V_{ij} \quad (2-6)$$

- Energie d'interaction noyaux-noyaux :

$$V_{nn} = \frac{1}{2} \sum_{\substack{\lambda,\beta=1 \\ \lambda \neq \beta}}^{N_n} \frac{Z_\lambda Z_\beta e^2}{|R_\lambda - R_\beta|} = \frac{1}{2} \sum_{\substack{\lambda,\beta=1 \\ \lambda \neq \beta}}^{N_n} V_{\lambda\beta} \quad \text{avec } Z_\lambda \text{ la charge d'un noyau} \quad (2-7)$$

- Energie d'interaction noyaux-électrons :

$$V_{ne} = - \sum_{i=1}^{N_e} \sum_{\lambda=1}^{N_n} \frac{Z_\lambda e^2}{|r_i - R_\lambda|} = \sum_{i=1}^{N_e} \sum_{\lambda=1}^{N_n} V_{i\lambda} \quad (2-8)$$

L'équation de Schrödinger peut s'écrire :

$$(T_n + V_{nn} + V_{ne} + V_{ee} + T_e)\psi(r_1, r_2, \dots, R_1, R_2, \dots) = E\psi(r_1, r_2, \dots, R_1, R_2, \dots) \quad (2-9)$$

Il existe des solutions analytiques de cette équation pour quelques systèmes très simples et des solutions numériques pour un nombre extrêmement réduit d'atomes. Cependant pour un nombre élevé de particules en interaction ($N_e + N_n$), un traitement exact s'est révélé une tâche impossible à réaliser. L'une des solutions préconisées est le recours à des approximations appropriées et simplificatrices visant à résoudre l'équation de Schrödinger dont nous allons citer un certain nombre.

2.2.2 L'approximation de Born - Oppenheimer

Cette approximation a été élaborée conjointement par *Born* et *Oppenheimer* en 1927 [4], elle consiste à prendre en compte l'énorme différence de masse qui existe entre les

noyaux et les électrons. Adoptée comme un premier niveau d'approche des problèmes à plusieurs corps en mettant en évidence le grand écart entre la masse des électrons du système (plus légers donc de plus grande mobilité) et celle des noyaux (relativement plus lourds $M \approx 1800m$ donc une mobilité plus réduite). Autrement dit, cette approximation est basée sur l'idée que les noyaux sont animés de mouvement relativement longs à celui des électrons de manière à les négliger sans grande erreur. Il ne reste donc à résoudre que l'hamiltonien électronique. Si l'on adopte l'hypothèse que les noyaux sont immobiles, l'équation de Schrödinger est considérablement simplifiée car le terme énergie cinétique des noyaux T_n est nul et le terme énergie d'interaction inter noyaux V_{nn} est réduit à une constante qui peut être rendue nulle par un choix convenable de l'origine de l'énergie potentielle. Nous pouvons définir une nouvelle fonction d'onde ψ_e des électrons et un nouvel hamiltonien électronique défini par les opérateurs.

$$H_e = T_e + V_{ne} + V_{ee} \quad (2-10)$$

ψ_e est solution de $H_e \psi_e = E_e \psi_e$

où

$$\left[\sum_{i=1}^{N_e} \frac{-\hbar^2}{2m} \Delta_i + \frac{1}{2} \sum_{\substack{i,j=1 \\ i \neq j}}^{N_e} \frac{e^2}{|r_i - r_j|} - \sum_{i=1}^{N_e} \sum_{\lambda=1}^{N_n} \frac{Z_\lambda e^2}{|r_i - R_\lambda^0|} \right] \psi(r, R_\lambda^0) = E_e(R_\lambda^0) \psi(r, R_\lambda^0) \quad (2-11)$$

Dans l'équation (2-11) le R_λ^0 figure non pas comme une variable mais plutôt comme un paramètre. L'énergie propre E_e est celle des électrons qui se déplacent dans un champ créé par les noyaux fixes. Bien que cette approximation permet de réduire de manière significative le nombre de variables nécessaires pour décrire la fonction ψ en éliminant tous les termes impliquant les noyaux uniquement, réduisant ainsi le degré de complexité à la résolution de l'équation de Schrödinger, « l'équation électronique » restant à résoudre demeure un problème à plusieurs corps. La nouvelle fonction d'onde totale du système dépend des coordonnées de tous les électrons et ne peut être découplée en contributions à une seule particule en raison de leur interaction et ne peut pas être résolue dans les calculs utilisant les ressources informatiques actuelles. En raison de cette difficulté, d'autres approximations supplémentaires sont requises pour réaliser de façon effective la résolution de l'équation de Schrödinger pour des matériaux réels.

2.2.3 Approximation de *Hartree*

Une solution approchée de l'équation (2-11) fut obtenue par *Hartree* en 1928 [5,6]. Cette approximation appelée aussi l'approximation self-consistant, consiste à réduire le problème de N_e corps à celui d'une seule particule, ce qui permet de considérer la fonction d'onde totale comme le produit de fonctions d'onde à une particule $\psi_i(r_i)$ orthogonale entre elles.

$$\psi(r) = \psi_1(r_1)\psi_2(r_2)\dots\psi_{N_e}(r_{N_e}) \quad (2-12)$$

Les électrons sont considérés comme indépendants, chacun d'eux se déplace dans un champ moyen créé par les autres électrons et les noyaux. L'équation de Schrödinger dite mono électronique où équation de *Hartree*, peut s'écrire comme suit :

$$h_i\psi_i(r) = \varepsilon_i\psi_i(r) \quad (2-13)$$

Où h_i est l'hamiltonien d'un électron. Qui s'écrit sous la forme, avec l'adoption des unités de *Hartree* ($\hbar = m = e = 1$)

$$h_i = -\frac{1}{2}\Delta_i + v_{ext}(r) + v_i(r) \quad (2-14)$$

Où v_{ext} représente le potentiel dû aux interactions entre les noyaux et les interactions entre électrons et les noyaux du système. Le potentiel $v_i(r)$ est le potentiel de *Hartree* pour le i^{eme} électron, il remplace son interaction électrostatique « électron-électron » avec tous les autres électrons du système, il s'exprime par

$$v_i(r) = \int d^3r' \frac{\rho_i(r')}{|r-r'|} \quad (2-15)$$

Où $\rho_i(r)$ est la densité électronique en r à laquelle contribuent tous les états monoélectroniques du système hormis celui sur lequel se trouve le i^{eme} électron, elle est sous la forme :

$$\rho_i(r) = \sum_{\substack{j=1 \\ j \neq i}}^{N_e} |\psi_j(r)|^2 \quad (2-16)$$

En substituant les équations (2-14), (2-15) et (2-16) dans (2-13) on obtient l'équation de *Hartree* pour un système monoélectronique :

$$\left(-\frac{1}{2}\Delta_i + v_{ext}(r)\right)\psi_i(r) + \sum_{\substack{j=1 \\ j \neq i}}^{N_e} \int d^3r' \frac{|\psi_j(r')|^2}{|r-r'|} \psi_i(r) = \varepsilon_i \psi_i(r) \quad (2-17)$$

Le potentiel de *Hartree* $v_i(r)$ qui détermine les fonctions d'onde monoélectroniques $\psi_i(r)$ est exprimé en termes de ces mêmes fonctions d'onde. C'est la raison pour laquelle cette approximation est appelée approximation du champ self-consistant. Cette méthode possède le mérite de proposer une solution self-consistante au problème de l'énergie d'un système électronique. La solution self-consistante reste toutefois une tâche très ardue, surtout si le nombre d'électrons N_e mis en jeu est grand. Souvent, on confond la densité monoélectronique avec la densité électronique totale.

$$\rho_i(r) = \rho(r) = \sum_{j=1}^{N_e} |\psi_j(r)|^2 \quad (2-18)$$

Ainsi le potentiel auquel est soumis chaque électron est le même, mais on introduit une interaction de chaque électron avec lui-même, ce qui est en toute rigueur incorrect surtout pour des systèmes localisés tels que les atomes.

2.2.4 Approximation de *Hartree-Fock*

En 1930, *Fock* [7] a montré que les solutions de l'hamiltonien (2-14) violent le principe de *Pauli* car non antisymétrique par rapport à l'échange de deux électrons quelconques. Le système électronique dans l'approximation de *Hartree* est incomplètement décrit. On peut présenter la différence entre l'énergie vraie du système multiélectronique et celle obtenue dans l'approche de *Hartree* comme celle qui représente le restant des interactions électroniques. Parmi les interactions manquantes figure l'énergie d'échange. C'est l'effet qui exprime l'antisymétrie de la fonction par rapport à l'échange des coordonnées de deux électrons quelconques du système. Une telle description obéit au principe d'exclusion de *Pauli* qui impose à deux électrons de même nombres quantiques de ne pouvoir occuper simultanément le même état quantique ainsi qu'au principe d'indiscernabilité. Dans la formulation de *Hartree* le principe d'indiscernabilité n'est pas respecté puisque l'électron i occupe précisément l'état i . En 1929 *Slater* montre qu'il est commode de représenter les fonctions d'onde antisymétriques sous une forme matricielle. Les fonctions d'onde respectant le principe de *Pauli* sont représentées par les déterminants de *Slater* [8]

$$\psi(r_1, \sigma_1, r_2, \sigma_2, \dots, r_{N_e}, \sigma_{N_e}) = \frac{1}{\sqrt{N_e!}} \begin{vmatrix} \psi_1(r_1, \sigma_1) & \psi_1(r_2, \sigma_2) & \dots & \psi_1(r_{N_e}, \sigma_{N_e}) \\ \psi_2(r_1, \sigma_1) & \psi_2(r_2, \sigma_2) & \dots & \psi_2(r_{N_e}, \sigma_{N_e}) \\ \dots & \dots & \dots & \dots \\ \psi_{N_e}(r_1, \sigma_1) & \psi_{N_e}(r_2, \sigma_2) & \dots & \psi_{N_e}(r_{N_e}, \sigma_{N_e}) \end{vmatrix} \quad (2-19)$$

Où σ représente le spin de l'électron

La fonction d'onde ψ_e donnée par l'équation (2-19) conduit aux équations de *Hartree-Fock* d'un système à une particule

$$\left(-\frac{1}{2} \Delta_i + v_{ext}(r) + \sum_{\substack{j=1 \\ j \neq i}}^{N_e} d^3 r' \frac{|\psi_j(r')|^2}{|r-r'|} \right) \psi_i(r) - \sum_{\substack{j=1 \\ j \neq i}}^{N_e} \delta_{\sigma_i, \sigma_j} \int d^3 r' \frac{\psi_j^*(r') \psi_i(r)}{|r-r'|} \psi_j(r) = \varepsilon_i \psi_i(r) \quad (2-20)$$

L'approximation de *Hartree Fock* [5], tient compte plus finement des interactions. En réalité les interactions électron-électron ne sont pas correctement prises en compte dans l'approche de *Hartree-Fock*. Les corrélations entre électron de spin antiparallèle, le caractère non local du potentiel d'échange sont autant d'éléments qui sont absent de l'équation (2-20). Malheureusement, le nombre de configuration augmente très rapidement avec le nombre d'électrons mis en jeu, et rend la résolution plus complexe ce qui limite la portée de ces calculs à des tous petits systèmes. Ces limitations ont été contournées en partie par la théorie de la Fonctionnelle de la densité (DFT), où c'est à partir de la densité électronique, et non des fonctions d'ondes, que l'équation de Schrödinger est résolue.

2.3 Principe de la théorie de la fonctionnelle de la densité (DFT)

2.3.1 Les théorèmes de *Hohenberg et Kohn*

La théorie de la fonctionnelle de la densité (*Density Functional Theory* :DFT) est l'une des méthodes les plus utilisées dans les calculs quantiques de la structure électronique de la matière (atomes, molécules, solides), aussi bien en physique de la matière condensée qu'en chimie quantique, cette méthode trouve ses origines dans le modèle développé par *L. Thomas* [9] et *E. Fermi* [10] à la fin des années 1920. Néanmoins il faudrait attendre le milieu des années 1960 et les contributions de *P. Hohenberg*, *Walter Kohn*, et *Lu Sham*. Alors que les méthodes *Hartree-Fock* conduisent à exprimer l'énergie du système comme une fonctionnelle de la fonction d'onde ψ , pour la méthode DFT l'énergie est une fonctionnelle de la densité

électronique ρ du système. Cela permet de reformuler le problème à la résolution de l'équation de *Schrödinger* électronique. Elle fournit une simplification conceptuelle considérable du problème étant donné qu'elle réduit le nombre de $3N$ variables pour N électrons à seulement 3 coordonnées spatiales (six si l'on considère les spin Up et Down pour le système magnétique). Cette reformulation de l'équation de *Schrödinger* basée sur la densité électronique $\rho(r)$ est la conséquence de deux théorèmes fondamentaux démontrés par *Hohenberg* et *Kohn* [1], que nous décrivons brièvement.

2.3.1.1 Premier théorème de *Hohenberg* et *Kohn*

Pour un ensemble de particules, le potentiel externe agissant sur ces particules détermine l'état fondamental de ce système et la densité de charge correspondante. Ainsi, toutes les quantités physiques concernant cet état (comme par exemple l'énergie totale du système) sont des fonctionnelles du potentiel externe.

Le premier théorème de *Hohenberg* et *Kohn* (1964) [1] assure que pour l'état fondamental ψ_0 , la relation entre la densité de l'état fondamentale ρ et le potentiel externe est bijective.

$$\rho(\vec{r}) \rightarrow \psi_0(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N) \rightarrow V_{ext}(\vec{r}) \quad (2-21)$$

Et en raison de cette correspondance biunivoque existant entre le potentiel externe \hat{V}_{ext} et la densité électronique de l'état fondamental $\rho_0(r)$, « *l'énergie totale du système à l'état fondamental est également une fonctionnelle unique universelle de la densité électronique* », soit :

$$E = E[\rho(r)] \quad (2-22)$$

Ce résultat constitue le premier théorème de *Hohenberg* et *Kohn*. Il est à la base de la théorie de la fonctionnelle de la densité. A la différence de la méthode *Hartree-Fock*, dans laquelle l'énergie totale du système est une fonctionnelle de la fonction d'onde, l'énergie totale du système à l'état fondamental est définie dans le formalisme de la DFT comme une fonctionnelle de la densité électronique de l'état fondamental. La démonstration du fait que l'énergie totale d'un système à l'état fondamental soit fonctionnelle de la densité électronique a permis à *Hohenberg-Kohn* d'exprimer cette fonctionnelle $E[\rho(r)]$ selon l'expression :

$$E[\rho(r)] = F_{HK}[\rho(r)] + \int \hat{V}_{ext}(r)\rho(r)dr \quad (2-23)$$

Dans laquelle $F_{HK}[\rho(r)]$ représente la fonctionnelle universelle de *Hohenberg-Kohn* et $\hat{V}_{ext}(r)$ représente le potentiel externe agissant sur ces particules.

2.3.1.2 Deuxième théorème de *Hohenberg et Kohn*

« Pour un potentiel \hat{V}_{ext} et un nombre d'électron N_e donnés, l'énergie totale du système atteint sa valeur minimale lorsque la densité $\rho(r)$ correspond à la densité exacte de l'état fondamental $\rho_0(r)$ ». (principe variationnel)

Ce qui peut être traduit par :

$$F_{HK}[\rho(r)] + \int \hat{V}_{ext}(r)\rho(r)dr = E_{EF} \quad (2-24)$$

Cette analyse a permis de souligner le fait que la connaissance de la fonctionnelle $F_{HK}[\rho(r)]$ suffirait à déterminer l'énergie totale du système ainsi que ses propriétés à l'état fondamental. Cependant, l'expression de cette fonctionnelle universelle n'est malheureusement pas connue. Il est par conséquent nécessaire de recourir à des approximations qui correspondent aux équations de *Kohn et Sham* établies dans l'objectif de fournir les fondements nécessaires pour exploiter de façon effective les théorèmes de *Hohenberg et Kohn* [1].

2.4 Les équations de *Kohn-Sham*

Thomas et Fermi, avaient tenté d'exprimer l'énergie totale d'un système en fonction de sa densité électronique, en représentant son énergie cinétique selon une fonctionnelle de cette grandeur. Mais les résultats obtenus ont donné une précision inférieure à celle de la méthode de *Hartree-Fock*, dans la mesure où le terme d'échange-corrélation n'était pas représenté. Ensuite *Dirac* a apporté une amélioration de cette théorie en ajoutant aux travaux de *Thomas et Fermi* une énergie d'échange fonctionnelle de la densité électronique. Cependant le terme de corrélation électronique demeurait toujours absent dans l'approche de *Thomas-Fermi-Dirac*. Et c'est finalement l'approche proposée par *Kohn-Sham* qui s'est imposée. En 1965, *Kohn et Sham* [2] ont eu l'idée de redéfinir le système d'électrons en interaction en un système d'électrons fictifs sans interaction et de même densité que celle caractérisant le système réel et évoluant dans un potentiel externe effectif $V_{eff}(r)$. Etant donné que le seul terme qu'elle laisse indéterminé constitue la plus faible contribution à l'énergie totale du système (le terme d'échange et de corrélation).

A ce stade d'approche, il est possible maintenant de déterminer la densité ρ et toutes les propriétés de l'état fondamental de tout système par une recherche de minimum de l'énergie, car celle-ci est fonctionnelle de la densité. L'énergie s'écrit :

$$E[\rho(r)] = F_{HK}[\rho(r)] + \int \hat{V}_{ext}(r)\rho(r)dr \quad (2-25)$$

L'expression de la fonctionnelle $F_{HK}[\rho]$ n'est malheureusement pas connue. Et le mieux que l'on puisse faire est de trouver une approximation de cette fonctionnelle. En général elle se décompose comme toute énergie à la somme de l'énergie cinétique $T[\rho]$ et en terme d'interaction électronique $W[\rho]$.

$$F_{HK}[\rho] = T[\rho] + W[\rho] \quad (2-26)$$

Le terme $W[\rho]$ peut s'écrire comme la sommation d'un terme de *Hartree* représentant l'énergie électrostatique classique d'une densité de charge ρ et des termes quantiques E_{xc}

$$W[\rho] = \frac{1}{2} \iint \frac{\rho(r)\rho(r')}{|r-r'|} d^3r d^3r' + E_{xc}[\rho] \quad (2-27)$$

Reste à déterminer les termes énergie cinétique $T[\rho]$ et la fonctionnelle $E_{xc}[\rho(r)]$ appelées « énergie d'échange - corrélation » qui rend compte des effets à plusieurs corps qui ne sont pas décrits dans les autres termes et contient toutes les différences entre le système fictif non interactif et le système réel interactif, incluant des corrections à la fois de l'interaction de Coulomb et de l'énergie cinétique proposé par *Kohn* et *Sham* qui consiste à remplacer le système d'électrons en interaction par un système d'électrons tous indépendants évoluant dans un potentiel externe effectif V_{eff} . Dont l'hamiltonien s'écrit :

$$H_s = T_s + V_{eff} \quad (2-28)$$

De manière analogue, on eut dire qu'il existe une fonctionnelle énergie totale de *Hohenberg* et *Kohn* d'écrite par :

$$E_s[\rho] = T_s[\rho] + \int v_{eff}(r)\rho(r)d^3r \quad (2-29)$$

Dont cette minimisation donne une densité exacte de l'état fondamental ρ_s . Le terme cinétique $T_s[\rho]$ est une fonctionnelle de la densité, représente l'énergie cinétique d'un système d'électrons sans interaction.

$$T_s[\rho] = \sum_i n_i \left\langle \phi_i \left| \left(-\frac{1}{2} \nabla_i^2 \right) \phi_i \right. \right\rangle \quad (2-30)$$

Où n_i est le nombre d'occupation de l'orbitale ϕ_i , avec $\sum_{i=1}^{N_{occ}} n_i = N_e$ (N_{occ} sont les états occupés).

$$\rho_s(r) = \sum_{i=1}^{N_{occ}} f_i |\phi_i(r)|^2 \quad \text{avec} \quad \sum_{i=1}^{N_{occ}} f_i = 1 \quad (2-31)$$

Où f_i qui peut être fractionnaire [11], est le nombre d'occupation de l'orbital i . Pour traiter les systèmes polarisés en spin, il suffit de séparer la sommation (2-31) en deux parties, l'occupation de chaque orbitale étant désormais fixée à 1 [12].

$$\rho(r) = \rho_\uparrow(r) + \rho_\downarrow(r) = \sum_{i=1}^{N_e^\uparrow} |\phi_i^\uparrow(r)|^2 + \sum_{i=1}^{N_e^\downarrow} |\phi_i^\downarrow(r)|^2 \quad (2-32)$$

Où ρ_\uparrow et ρ_\downarrow désignent respectivement les densités associées aux états de spin up (\uparrow) et down (\downarrow), avec $N_e^\uparrow + N_e^\downarrow = N_e$

L'hypothèse centrale de *Kohn* et *Sham* est que pour tout système en interaction on peut trouver un potentiel V_s tel que la densité exacte du système $\rho(r)$ soit égale à la densité du système d'électrons indépendant $\rho_s(r)$.

On a ainsi remplacé un système d'électrons en interaction par un système sans interaction, mais évoluant dans un potentiel effectif V_{eff} , formulé en tant que fonctionnelle de la densité électronique et qui s'écrit :

$$V_{eff}(r) = V_{eff}[\rho(r)] = V_{ext}(r) + \frac{\delta E_H[\rho(r)]}{\delta \rho(r)} + \frac{\delta E_{xc}[\rho(r)]}{\delta \rho(r)} = V_{ext}(r) + \int \frac{\rho(r')}{|r-r'|} dr + V_{xc}(r) \quad (2-33)$$

où $V_{xc}(r)$ est le potentiel d'échange-corrélation est déterminé grâce au potentiel associé

$$V_{xc}(r) = \frac{\delta E_{xc}[\rho(r)]}{\delta \rho(r)} \quad (2-34)$$

L'équation de *Schrödinger* à résoudre dans le cadre de l'approche de *Kohn-Sham* est ainsi sous la forme :

$$\left[\underbrace{-\frac{1}{2} \nabla_i^2 + V_{eff}(r)}_{H_{KS}} \right] \phi_i(r) = \epsilon_i \phi_i(r) \quad (2-35)$$

Où l'opérateur H_{KS} est appelé hamiltonien de *Kohn* et *Sham*.

Les équations (2-32, 2-34, 2-35) correspondent aux équations de *Kohn-Sham*, elles doivent être résolues de manière self-consistante (en débutant à partir d'une certaine densité initiale) pour trouver la densité d'état fondamental (voir figure 2-1). Tous les calculs de type DFT sont basés sur la résolution itérative de ces trois équations. Enfin l'élaboration des équations de *Kohn et Sham* a permis de mettre en évidence le fait que la seule fonctionnelle de la densité demeurant inconnue au sein de ce formalisme correspond à la fonctionnelle d'échange-corrélation $E_{xc}[\rho(r)]$. Afin de résoudre les équations de *Kohn-Sham*, diverses fonctionnelles d'échange-corrélation ont de ce fait été envisagées.

2.5 La fonctionnelle d'échange et corrélation

2.5.1 L'approximation de la densité locale (LDA)

Historiquement, l'approximation de la densité locale (LDA : *Local Density Approximation*) [13] est la première fonctionnelle proposée par *Kohn et Sham*, qui ont remarqué que les solides peuvent souvent être considérés comme proches de la limite du gaz d'électrons homogène pour lequel les effets d'échange et de corrélation sont locaux. L'idée de l'approximation de la densité locale définie en un point r , dépend faiblement des variations de la densité autour de ce point. Ainsi, en un point r auquel correspond une densité $\rho(r)$, il sera associé un potentiel d'échange et corrélation comparable à un gaz d'électrons de même densité $\rho(r)$. L'énergie d'échange et corrélation totale E_{XC} s'exprime alors en fonction de l'énergie d'échange et corrélation par particule ε_{xc} par l'expression :

$$E_{XC}^{LDA}[\rho(r)] = \int \varepsilon_{xc}^{LDA}[\rho(r)]\rho(r)d^3r \quad (2-36)$$

Le potentiel d'échange-corrélation peut être obtenu d'une façon variationnelle selon l'équation :

$$V_{XC}^{LDA}(r) = \frac{\delta(\rho(r)\varepsilon_{xc}^{LDA}[\rho(r)])}{\delta\rho(r)} \quad (2-37)$$

La fonctionnelle d'échange-corrélation peut être divisée en une contribution d'échange et une de corrélation.

$$\varepsilon_{xc}^{LDA}[\rho(r)] = \varepsilon_x[\rho(r)] + \varepsilon_c[\rho(r)] \quad (2-38)$$

La contribution d'échange est déterminée analytiquement pour un gaz d'électrons homogène telle que calculée par Dirac [14].

$$\varepsilon_x^{\text{hom}}[\rho(r)] = -C_x \rho^{1/3}(r) \quad (2-39)$$

$$\text{Avec : } C_x = \frac{3}{4} \left(\frac{3}{\pi} \right)^{1/3}$$

La relation (2-39) peut être légèrement reformulée à partir de la densité totale de la polarisation de spin, l'approximation s'appelle alors LSDA qui s'écrit [8] :

$$\varepsilon_x[\rho(r)] = -\frac{1}{2} C_x \rho^{1/3} \left[(1 + \zeta)^{3/4} + (1 - \zeta)^{4/3} \right] \quad (2-40)$$

Dans laquelle $\zeta = \frac{\rho^\uparrow - \rho^\downarrow}{\rho^\uparrow + \rho^\downarrow}$ est la polarisation de spin (différence normalisée entre ρ^\uparrow et ρ^\downarrow).

Enfin, concernant la contribution de corrélation, *Ceperly* et *Alder* [15] et plus récemment *Ortiz* et *Ballone* [16] ont déterminé numériquement cette contribution par la simulation du type Monté-Carlo quantique. Il existe ainsi de nombreuses paramétrisations pour la contribution de corrélation dont celles de *Hedin-Lundqvist* [17] et *Vosko-Wilk-Nusair* [18]. L'approximation LDA, bien que rudimentaire dans sa conception permet d'obtenir de bons résultats. Elle permet de décrire précisément des propriétés comme la distance d'équilibre, le module de compressibilité. De façon plus générale, la LDA permet de décrire les liaisons covalentes, mais ne peut pas en principe traiter les systèmes dans lesquels les corrélations non locales sont importantes. Ce qui est typiquement le cas pour les interactions de *Van der Walls*. La LDA sous-estime en général légèrement les distances de liaisons et surestime les énergies de cohésion. Ainsi différentes améliorations ont été proposées. Une première extension consiste à prendre en compte le spin dans la fonctionnelle d'échange-corrélation, l'ajout de ce degré de liberté permet en effet une plus grande flexibilité qui peut améliorer les approximations pour les systèmes polarisés. Une autre démarche consiste à aller au-delà de l'approximation locale.

2.5.2 L'approximation de gradient généralisé (GGA)

Dans le but, d'aller au-delà de la vision locale de la densité LDA pour des systèmes inhomogènes en tenant compte de la variation locale de la densité $\rho(r)$ à travers son gradient $|\nabla\rho(r)|$. La LDA peut alors être réinterprétée comme le premier terme d'un développement en série de Taylor de l'énergie d'échange et de corrélation en terme de la densité électronique $\rho(r)$. Cette approche appelée approximation de développement de gradient (GEA : Gradient Expansion Approximation) aurait dû améliorer les résultats apportés

par la LDA. En réalité sa mise en œuvre a donné des résultats sensiblement moins bons que ceux de la LDA. La raison en est que ce développement a fait perdre toute signification physique au trou d'échange et de corrélation [19]. Ces problèmes ont été contournés et la fonctionnelle a été modifiée de façon à respecter les principales conditions aux limites et les fonctionnelles qui ont résulté ont été appelées approximation de gradient généralisé (GGA : Generalized Gradient Approximation). Ces fonctionnelles de type GGA sont à l'origine de l'utilisation massive de la DFT au sein de la communauté des physiciens, l'utilisation d'une fonctionnelle GGA a permis d'accroître de façon significative la précision des calculs comparativement à la description fournie par la LDA. La forme générale de l'approximation de la GGA est donnée par :

$$E_{xc}^{GGA}[\rho(r)] = \int f_{xc}^{GGA}[\rho(r), \nabla\rho(r)] d^3r \quad (2-41)$$

Ou encore

$$E_{xc}^{GGA}[\rho_{\uparrow}, \rho_{\downarrow}] = \int f_{xc}^{GGA}[\rho_{\uparrow}(r), \rho_{\downarrow}(r), \nabla\rho_{\uparrow}(r), \nabla\rho_{\downarrow}(r)] d^3r \quad (2-42)$$

Où la fonctionnelle f_{xc}^{GGA} dépend de la forme de la GGA utilisée. Les fonctionnelles GGA traitent en général séparément la partie échange et la partie corrélation. Pour la convenance, nous donnons seulement la partie d'échange (qui représente la plus grande partie de l'énergie d'échange et de corrélation) des fonctionnelles. La fonction d'énergie d'échange de la GGA s'écrit sous la forme

$$E_{xc}^{GGA}[\rho] = -\frac{3}{4} \left(\frac{3}{\pi} \right)^{1/3} \int \rho^{4/3} F[s(r)] d^3r \quad (2-43)$$

Où $F(s)$ est le facteur de perfectionnement avec $s = \frac{|\nabla\rho|}{2\rho(3\pi^2\rho)^{1/3}}$

Dans le cas de la LDA $F(s) = 1$ est la fonctionnelle exacte pour le gaz homogène d'électrons, qui est différent de 1, pour un développement de la GGA.

Perdew et *Wang* [20] ont exprimé analytiquement le facteur $F(s)$ de la partie d'échange

$$F^{PW91}(s) = \frac{1 + 0.1645s + \sinh^{-1}(7.7956s) + (0.2743 - 0.1508e^{-100s^2})s^2}{1 + 0.1645s + \sinh^{-1}(7.795s) + 0.004s^4} \quad (2-44)$$

Ensuite *Perdew*, *Burke*, et *Ernzerhof* (PBE) [21] ont développé une nouvelle forme analytique du facteur $F(s)$ défini comme suit :

$$F^{PBE}(s) = 1 + \kappa - \frac{\kappa}{1 + \frac{\vartheta}{\kappa} s^2} \quad (2-45)$$

avec $\kappa = 0.804$

$$\vartheta = 0.21951$$

récemment *Z.Wu* et *R.E.Cohen* [22] ont proposé une forme du facteur $F(s)$ donnée par :

$$F^{PBE}(s) = 1 + \kappa - \frac{\kappa}{1 + \frac{x}{\kappa} s^2} \quad (2-46)$$

avec $\kappa = 0.804$

$$x = \frac{10}{80} s^2 + \left(\vartheta - \frac{10}{80}\right) s^2 e^{-s^2} + \ln(1 + cs^4)$$

$$\vartheta = 0.21951, \text{ et } c = 0.0079325$$

Plusieurs expressions ont été proposées pour ce facteur conduisant à autant de fonctionnelles, mais la fonctionnelle PBE est l'une des plus utilisées en physique et donne des résultats tout à fait satisfaisants.

2.6 Méthodes de calcul des structures électroniques

Nous allons décrire schématiquement les grandes étapes d'un calcul DFT qui consiste à résoudre les équations de *Kohn-Sham* de manière itérative (figure2-1). Pour cela en première étape on impose une densité électronique d'entrée au pas numéro i , $\rho_i^{\text{entrée}}(r)$ dont on déduit un potentiel effectif d'entrée. La solution des équations de *Kohn-Sham* produit un ensemble de N énergies associées à N fonction d'ondes. Le problème aux valeurs propres de l'équation (2-35) est ainsi résolu. Une densité électronique résultante est construite à partir des vecteurs propres en utilisant l'équation (2-31), cette nouvelle densité $\rho_{i+1}^{\text{entrée}}(r)$ est alors « mixée » à la densité initialement introduite pour le calcul, afin d'obtenir des données affinées pour l'itération suivante. Dans le cas le plus général, c'est une combinaison linéaire entre $\rho_i^{\text{entrée}}(r)$ et $\rho_i^{\text{sortie}}(r)$.

$$\rho_{i+1}^{\text{entrée}}(r) = (1 - \alpha)\rho_i^{\text{entrée}}(r) + \alpha\rho_i^{\text{sortie}}(r) \quad (2-47)$$

Avec i l'exposant indiquant le rang de l'itération, α un paramètre constant à chaque itération ($0 < \alpha \leq 1$).

La procédure des itérations est poursuivie jusqu'à ce que le critère de convergence soit atteint. En pratique, on s'attache à ce que l'énergie totale ne varie plus à un seuil de tolérance près et on appelle ceci atteindre l'auto cohérence.

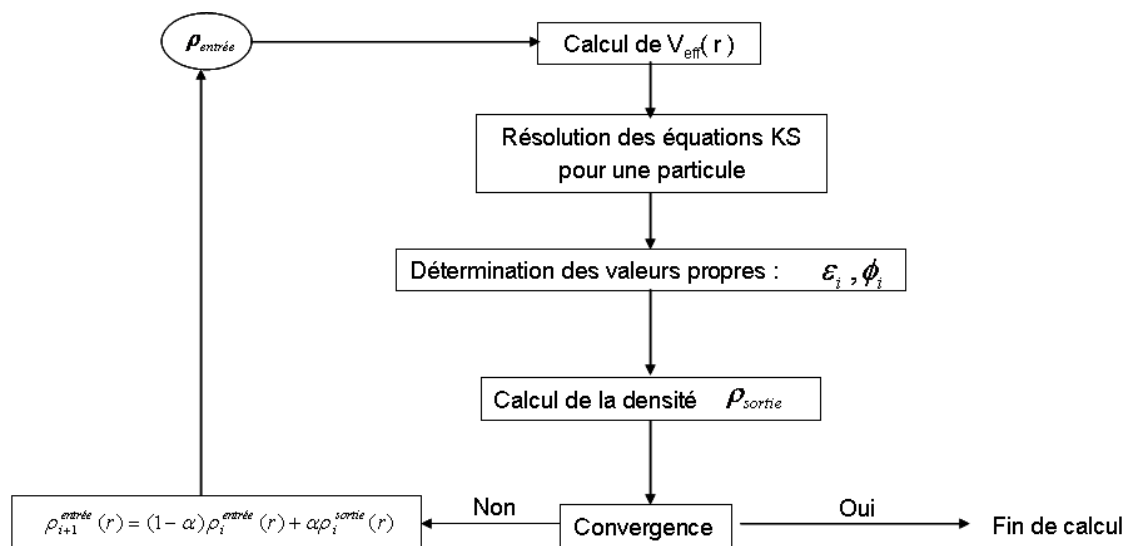


Figure 2-1 : Organigramme d'un calcul auto-cohérent dans une méthode basée sur la théorie de la fonctionnelle de la densité électronique.

2.6.1 Choix de la méthode, de la base des fonctions d'ondes et de la forme du potentiel

Plusieurs méthodes de calcul de structure électronique existent. Leur point commun est la résolution des trois équations de *Kohn-Sham* de façon auto cohérente. Leurs spécificités respectives est de représenter, de décrire le potentiel de la densité électronique et les orbitales monoatomiques de *Kohn-Sham*. Le choix de la méthode doit apporter un niveau de précision élevé sur les résultats obtenus ainsi minimiser le coût en temps de calcul.

La figure (2-2) donne un aperçu sur les différents traitements envisageables pour les termes de l'équation de *Kohn-Sham* [23]. Le terme d'échange et de corrélation a été séparé du potentiel effectif afin d'illustrer la manière dont les différentes méthodes d'approximations. Les effets relativistes des électrons indépendants peuvent être inclus dans le terme d'énergie cinétique. La périodicité ou son absence du composé étudié peut être prise en compte à travers la construction du potentiel, sa forme peut être plus ou moins précise selon un potentiel « tous électrons » ou « pseudopotentiel ». Les états des spin des électrons peuvent être réalisés. Finalement, la base utilisée pour représenter les orbitales de *Kohn* et *Sham* peut être très variée. Elle peut être constituée de fonctions localisées ou non, ou entièrement numériques. Dans ce travail, nous allons présenter principalement l'utilisation des ondes planes comme base d'expansion de la fonction d'onde.

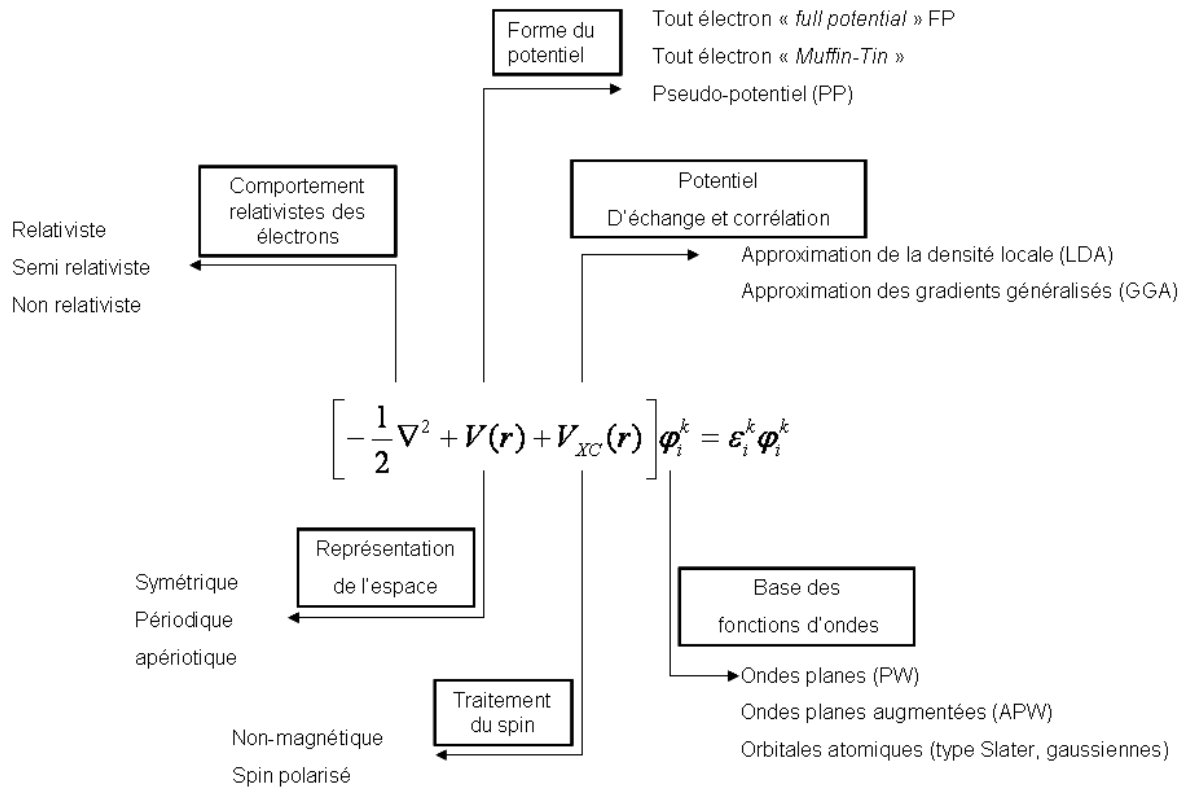


Figure 2-2 : Les différents traitements de l'énergie cinétique, du potentiel, les fonctions d'ondes.

2.6.2 Théorème de Bloch

Le traitement de *Kohn* et *Sham* nous à ramené à la résolution de trois équations interdépendantes qui sont, le potentiel effectif (2-34.), les fonctions d'onde monoélectroniques (2-35), et la densité électronique (2-32). La résolution de ses trois équations nécessite en général l'utilisation de base afin de représenter la densité électronique $\rho(r)$, le potentiel $V(r)$, et en particulier les orbitales de *Kohn* et *Sham* $\phi_i(r)$.

Pour un système périodique, en vertu du théorème de *Bloch*, qui stipule « toute fonction d'onde monoélectronique d'un réseau cristallin $\phi_k^n(\vec{r})$, peut être exprimée comme produit d'une onde plane $e^{(i\vec{k} \cdot \vec{r})}$ pour une fonction de même périodicité que le potentiel $u_k^n(\vec{r})$ ».

$$\phi_k^n(\vec{r}) = e^{(i\vec{k} \cdot \vec{r})} u_k^n(\vec{r}) \tag{2-48}$$

Avec $u_k^n(\vec{r} + \vec{R}) = u_k^n(\vec{r})$

\vec{k} : Vecteur d'onde de la première zone de Brillouin du réseau réciproque du cristal, et n correspond à l'indice de bande.

\vec{R} : Vecteur de translation du réseau direct.

Soulignons que l'emploi du théorème de *Bloch*, permet d'effectuer les calculs dans une cellule elle-même partie du réseau dans l'espace réel. Le réseau réciproque associé est également périodique et sa cellule élémentaire est appelée première zone de Brillouin. Chaque reproduction de la première zone de Brillouin est une zone de Brillouin.

2.6.3 La méthode d'ondes planes

L'idée la plus simple pour représenter les fonctions d'onde $\phi_k^n(\vec{r})$ est un développement en ondes planes, à l'aide de séries de Fourier.

$$\phi_k^n(\vec{r}) = \Omega^{-1/2} \sum_{\vec{K}} C_{\vec{K}}^{n,\vec{k}} e^{i(\vec{k}+\vec{K})\cdot\vec{r}} \quad \text{Avec } n=1,\dots,N_e \quad (2-49)$$

Où Ω représente le volume de la cellule de simulation, \vec{K} et \vec{k} représentent respectivement le vecteur du réseau réciproque et le vecteur d'onde de l'espace réciproque appartenant à la première zone de Brillouin.

Avec une telle décomposition, les équations de *Kohn-Sham* peuvent être théoriquement résolues. Mais en pratique deux considérations font obstacle aux calculs. D'une part il existe une infinité de vecteur \vec{K} dans le réseau réciproque, d'autre part les vecteurs \vec{k} qui appartiennent à la première zone de Brillouin sont également en nombre infini. Le premier obstacle peut être levé en définissant une énergie de coupure E_c au delà de laquelle les vecteurs \vec{K} ne sont pas pris en considération qui correspond à un échantillonnage dans l'espace réel des fonctions d'onde et qui revient à sélectionner une sphère de rayon K_{\max} dans le réseau réciproque. Et pour les vecteurs \vec{k} , on discrétise la première zone de Brillouin et on suppose qu'il existe une évolution continue des bandes entre deux points \vec{k} . De nombreuses procédures existent pour générer les pavages de \vec{k} , nous citerons celle de *Chadi* et *Cohen* [24] et son extension par *Monkhorst* et *Pack* [25] et qui a été utilisée dans notre travail.

2.6.4 Ondes planes augmentées (APW) et ondes planes augmentées linéarisées (LAPW)

2.6.4.1 Méthode des ondes planes augmentées (APW)

L'élaboration de la méthode APW, est basée sur l'observation de *Slater* [26] qui tient compte des électrons au voisinage du noyau qui sont localisés et les électrons entre les atomes

qui sont délocalisés. De ce fait on développe les fonctions d'ondes pour les électrons localisés non pas par une base d'ondes planes mais par une base combinant des orbitales atomiques localisées au voisinage des noyaux et une base d'ondes planes pour les électrons entre les atomes. Par conséquent l'espace peut être divisé en deux régions (figure 2-3).

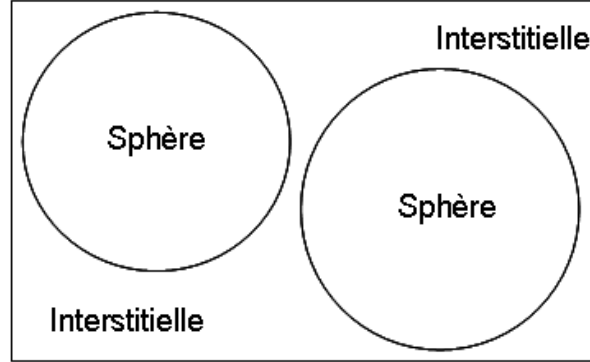


Figure 2-3 : Représentation de la cellule unitaire, la région interstitielle et la région sphérique.

- ✚ Région I : des sphères appelées « Muffin-Tin » englobant des atomes, dont les fonctions d'ondes sont des fonctions radiales multipliées par des harmoniques sphériques dans les sphères atomiques « Muffin-Tin ».
- ✚ Région II : dite région interstitielle délimitant l'espace résiduel non occupé par des sphères, dont les fonctions d'ondes sont développées par une base d'ondes planes.

Les deux régions sphériques et interstitielles sont définies par les fonctions d'ondes :

$$\phi_{\vec{k}}(\vec{r}, E) = \begin{cases} \frac{1}{\Omega^{1/2}} \sum_{\vec{k}} C_{\vec{k}} e^{i(\vec{k}+\vec{k})\cdot\vec{r}} & r < R_{\alpha} \\ \sum_{l,m} A_{l,m}^{\alpha} u_l^{\alpha}(\vec{r}, E) Y_l^m(\hat{r}) & r > R_{\alpha} \end{cases} \quad (2-50)$$

où Ω est le volume de la cellule unitaire de simulation (ou la cellule de base du réseau) :

u_l^{α} : des solutions régulières de l'équation (2-51)

$Y_{l,m}(\hat{r})$: les harmoniques sphériques

$A_{l,m}^{\alpha}$: les coefficients du développement en harmoniques sphériques

R_{α} : le rayon de la sphère « Muffin-Tin ».

$$\left\{ -\frac{d^2}{dr^2} + \frac{l(l+1)}{r^2} + V^{\alpha}(r) - E_l \right\} r u_l^{\alpha}(r, E) = 0 \quad (2-51)$$

Slater a justifié que les ondes planes sont des solutions de l'équation de Schrödinger lorsque le potentiel est constant et les fonctions radiales sont aussi solutions de l'équation de Schrödinger dans un potentiel sphérique.

Les coefficients du développement en harmoniques sphériques $A_{l,m}^\alpha$ donnés par l'équation (2-52), peuvent être déterminés en imposant que les fonctions d'ondes doivent être continues aux bords des sphères.

$$A_{l,m}^\alpha = \frac{4\pi i^l}{\Omega^{1/2} u_l^\alpha(R^\alpha)} \sum_{\vec{k}} C_{\vec{k}} j_l(|\vec{k} + \vec{K}| R^\alpha) Y_{l,m}^*(\vec{k} + \vec{K}) \quad (2-52)$$

Dans la méthode APW, on distingue quelques difficultés de calcul, dont celle liée au problème de l'asymptote, de la présence au dénominateur du terme $u_l^\alpha(R^\alpha)$, car il est donc possible de trouver des valeurs de l'énergie pour lesquelles $u_l^\alpha(R^\alpha)$ s'annule à la limite de la sphère et par conséquent, les coefficients du développement en harmoniques sphériques $A_{l,m}^\alpha$ divergent. Afin de surmonter les difficultés, plusieurs améliorations ont été apportées à cette méthode, dont les fonctions radiales multiples sont choisies pour avoir la même dérivée logarithmique et adaptées à des ondes planes à condition que la valeur de la fonction et sa première dérivée radiale soient continues.

2.6.4.2 Méthode linéaire des ondes planes augmentées (LAPW)

En 1975, *Andersen* a proposé une méthode dans laquelle les fonctions de base et leurs dérivées sont continues pour une énergie fixe. Ce choix a permis de résoudre les problèmes rencontrés par APW, en donnant ainsi une méthode de structure des bandes, flexible et exacte, cette méthode est appelée la méthode linéaire des ondes planes augmentées (LAPW). Ou les fonctions de base à l'intérieur de la sphère *Muffin-Tin* sont une combinaison linéaire des fonctions radiales $U_l(r)Y_{lm}(\hat{r})$ et de leurs dérivées par rapport à l'énergie $\dot{u}_l(r)Y_{lm}(\hat{r})$.

Les $u_l(r)$ sont définies comme dans la méthode APW avec une énergie fixe, les fonctions de base son alors données par :

$$\phi_{\vec{k}}^{\vec{k}}(\vec{r}, E) = \begin{cases} \frac{1}{\Omega^{1/2}} \sum_{\vec{k}} C_{\vec{k}} e^{i(\vec{k} + \vec{K}) \cdot \vec{r}} & r > R_\alpha \\ \sum_{l,m} [A_{l,m}^\alpha u_l(r) + B_{l,m}^\alpha \dot{u}_l] Y_{l,m}(r) & r < R_\alpha \end{cases} \quad (2-53)$$

B_{lm}^α sont des coefficients correspondants à la fonction u_l^α et sont de même nature que les coefficients A_{lm}^α . Ils sont déterminés de telle sorte à satisfaire aux conditions de continuité entre les deux régions. Ainsi la synthèse des fonctions de base LAPW consiste en premier à la détermination des fonctions radiales et leurs dérivées, ensuite déterminer les coefficients A_{lm}^α et B_{lm}^α qui satisfont les conditions aux limites de la sphère Muffin -Tin. La dépendance énergétique des fonctions d'onde a été levée. De plus une valeur nulle de u_l en R_α est sans conséquence sur la continuité à la surface de la sphère, car sa dérivée sera différente de zéro, LAPW fournit une base suffisamment flexible pour décrire correctement les fonctions propres avec énergie propres près de l'énergie de linéarisation, qui peut être gardé fixe, et contrairement à APW, cette méthode nous permet d'obtenir toutes les énergies propres à travers une seule diagonalisation. Elle permet également d'obtenir avec une seule énergie E_l , toutes les bandes de valence dans un grand intervalle d'énergie. Mais, avec cette procédure, les calculs perdent en précision par rapport à la méthode APW qui reproduit les fonctions d'onde très correctement, tandis que LAPW entraîne une erreur sur les fonctions d'onde de l'ordre de $(E - E_l)^2$ et une autre sur les énergies de bandes de l'ordre de $(E - E_l)^4$. Pour plus de détails sur les méthodes APW et LAPW, on peut consulter les références [19, 27, 28].

2.7 La méthode des pseudopotentiels

L'approche de la méthode pseudopotentiel repose sur l'hypothèse que les propriétés physiques et chimiques des matériaux sont essentiellement gouvernées par des électrons de valence. En effet dans la plupart des cas les électrons de périphérique (de valence) contribuent seuls à la formation des liaisons chimiques. Les électrons de cœur pourront être regroupés avec les noyaux pour former des ions rigides : c'est l'approximation du cœur rigide. Ainsi seuls les électrons de valence (les plus éloignés du noyau) seront pris en compte pour les liaisons chimiques et afin de tenir compte des interactions qui ont perdu leur caractère explicite, le potentiel effectif dans les équations de *Kohn-Sham* doit être remplacé par un pseudopotentiel incluant ; l'interaction du noyau avec les électrons de cœur; le potentiel de *Hartree* des électrons de cœur et une composante d'échange et de corrélation due à l'interaction entre électrons de cœur et de valence.

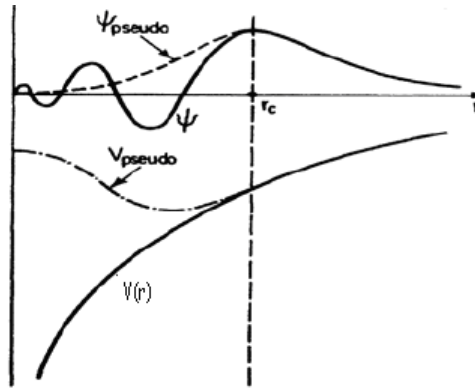


Figure 2-4 : Représentation schématique du remplacement d'une fonction d'onde φ et du potentiel $V(r)$ par une pseudo fonction d'onde φ_{pseudo} et un pseudo potentiel V_{pseudo} respectivement.

Ce pseudopotentiel tente en fait à reproduire l'interaction générée par le vrai potentiel sur les électrons de valence sans inclure explicitement dans le calcul les électrons de cœur, les fonctions d'onde relativement oscillantes dans la région du cœur (principe de répulsion de Pauli), et qui seraient relativement difficile à décrire à partir d'une base d'ondes planes (nombre de vecteurs \vec{K} très élevé), sont remplacées par des pseudo fonctions d'onde qui sont dépourvues de nœuds dans la région du cœur (figure 2-4). Le pseudopotential construit de manière à ce que les caractéristiques de déphasage qu'il produit sur les pseudo fonctions d'onde soient identiques à celles résultantes de l'action du cœur ionique sur les vraies fonctions d'onde de valence, tout en générant des pseudo fonctions d'onde dépourvues d'oscillation dans la région du cœur ionique. Au delà de cette région du cœur, délimitée par r_c , l'égalité $\varphi_i^{pseudo}(r) = \varphi_i(r)$ est imposée, (les pseudo fonctions d'onde doivent être identiques aux fonctions d'onde de valence vraies), qui sont appelées, fonctions d'onde de valence « tous électrons ». Ces pseudo-fonctions d'onde sont plus lisses que les vraies fonctions d'onde et elles peuvent de ce fait être représentées correctement en utilisant un nombre de vecteurs \vec{K} très inférieur à celui nécessité dans le traitement des vraies fonctions d'onde.

La construction d'un pseudopotential doit être guidée par la volonté d'obtenir les mêmes énergies propres pour les atomes qu'avec un calcul incluant tous les électrons. Par ailleurs, il ne doit pas diverger au voisinage du noyau afin de ne pas faire apparaître de terme de fréquence élevée dans l'espace réciproque [29]. Hamman, Schluter et Chiang [30] et Kerker [31] ont mis au point, dans ce contexte, une méthode qui garantit une description de la densité de charge de valence. Celle-ci consiste à conserver la norme : les pseudo-fonctions

d'onde de valence sont normées à l'unité. Les pseudopotentiels associés ont été appelés pseudopotentiels à norme conservée.

Toutefois, l'utilisation des pseudopotentiels à normes conservées est très coûteuse dans le cas des métaux de transition. Ces atomes possèdent des orbitales localisées et nécessitent un nombre d'ondes planes important. D'autres classes de pseudopotentiels ont été proposées pour contourner la contrainte de conservation de la norme, ce sont les pseudopotentiels à norme non conservée ou relaxée. A cet effet *Vanderbilt* [32] a eu l'idée de relâcher la contrainte de la conservation de la norme, en construisant une nouvelle classe de pseudopotentiels dans lesquels les pseudo-fonctions d'onde peuvent être arbitrairement lisses dans la région du cœur. Ils sont qualifiés « ultrasoft » puis généralisés par *Blöchl* [33] pour l'approche des ondes augmentées de projecteurs, PAW (Projector Augmented Wave) que nous décrivons dans la section suivante.

2.8. Le formalisme PAW

L'idée de la méthode PAW est née du souci de restreindre le grand nombre de fonctions de base pour décrire une fonction d'onde typique dans un solide. L'approche PAW consiste à utiliser deux sortes de fonctions de base : l'une utilise des fonctions d'onde atomiques partielles à l'intérieur des cœurs, région où il y a beaucoup d'électrons dans un faible volume. Les fonctions d'ondes oscillent rapidement entraînant de fortes variations d'amplitude de la densité de charge électronique totale, l'autre est un ensemble de fonctions d'ondes planes dans la région interstitielle, où la densité électronique totale est faible, par conséquent, quelques ondes planes suffisent à décrire les états de valence de manière satisfaisante. Il y a donc séparation de l'espace en deux zones [34].

L'objectif de la méthode PAW est de déterminer un ensemble de fonctions de base et de projecteurs pour relier la fonction d'onde de valence $|\psi\rangle$ à une pseudo-fonction d'onde de valence $|\tilde{\psi}\rangle$ par l'équation dans une représentation de *Dirac* qui est extrêmement pratique pour ce type de raisonnement [33,35].

$$|\psi\rangle = |\tilde{\psi}\rangle + \sum_j \left(|\phi_j\rangle - |\tilde{\phi}_j\rangle \right) \langle \tilde{P}_j | \tilde{\psi}\rangle \quad (2-54)$$

Avec $|\phi_j\rangle$ un état fictif de valence, $|\tilde{\phi}_j\rangle$ un pseudo-état fictif de valence et $|\tilde{P}_j\rangle$ un projecteur. Dans l'équation j représente à la fois l'atome considéré, le moment angulaire $L_j = \{l_j, m_{l_j}\}$ (avec l_j le nombre quantique de moment cinétique orbital et m_{l_j} le nombre quantique

magnétique). La synthèse d'un pseudopotentiel débute toujours par un calcul tous électrons sur l'atome isolé, dans une configuration de référence et pour une approximation définie de la fonctionnelle d'échange-corrélation.

Le calcul des fonctions de base et des projecteurs introduits dans l'équation (2-54) nécessite quatre étapes.

✚ On calcule les fonctions de base $|\varphi_j\rangle$ par résolution auto-cohérente des équations de *Kohn-Sham* pour l'atome considéré.

$$H(r)|\varphi_j\rangle = \varepsilon_j|\varphi_j\rangle \quad (2-55)$$

La densité électronique totale de l'atome, $\rho(r)$ peut être partitionnée entre une densité électronique de cœur $\rho_c(r)$ et une densité électronique de valence $\rho_v(r)$ et en appliquant l'approximation des électrons de cœur fixes '*frozen core approximation*', l'hamiltonien de *Kohn-Sham* s'écrit :

$$H(r) = -\frac{1}{2}\nabla_r^2 + V_{eff}(r) \quad (2-56)$$

Le potentiel effectif a pour expression :

$$V_{eff}(r) = V_{ext}(r) + \int \frac{n_v(r') + n_c(r')}{|r - r'|} d^3r' + V_{xc}[n_c(r) + n_v(r)] \quad (2-57)$$

Comme l'équation (2-33), le premier terme du membre de droite de l'équation est le potentiel externe imposé par le noyau, le second terme est l'interaction coulombienne classique de *Hartree* et le troisième est le potentiel d'échange et de corrélation.

On choisit un rayon r_c définissant une sphère d'augmentation centrée sur l'atome, dont il ne doit pas être trop important pour que les sphères d'augmentation des différents atomes ne se recouvrent pas. Cependant il ne doit pas être trop faible pour que la densité de cœur $\rho_c(r)$ soit effectivement contenue dans cette sphère.

✚ On calcule ensuite les pseudo-fonctions de base $|\tilde{\varphi}_j\rangle$ en résolvant les équations de *Kohn-Sham* pour un pseudo-hamiltonien $\tilde{H}(r)$:

$$(\tilde{H}(r) - \varepsilon_j)|\tilde{\varphi}_j\rangle = C_j g(r)|\tilde{\varphi}_j\rangle \quad (2-58)$$

Avec ε_j les valeurs propres obtenues après résolution des équations (2-55). C_i un paramètre variable qui permet d'une part d'ajuster le nombre de nœuds (nombre de valeurs de r en

dehors de zéro pour lesquelles la partie radiale de la fonction d'onde s'annule) de $|\tilde{\varphi}_j\rangle$ pour chaque moment angulaire et d'autre part d'assurer qu'en dehors de la sphère d'augmentation :

$$|r| > r_c \Rightarrow |\tilde{\varphi}_j\rangle = |\varphi_j\rangle \quad (2-59)$$

En pratique cette condition est obtenue en modifiant la valeur de C_j pour que les dérivées logarithmiques de $|\tilde{\varphi}_j\rangle$ et $|\varphi_j\rangle$ soient égales.

$g(r)$ une fonction de forme qui s'annule en dehors de la région d'augmentation qui peut prendre deux formes différentes, celles proposée par *Bloch* [33] et *Holzwarth et coll* [35];

$$g(r) = e^{-(r/r_c)^6} \quad (2-60)$$

et

$$g(r) = \begin{cases} \left[\frac{\sin(\pi r / r_c)}{(\pi r / r_c)} \right] & \text{si } |r| < r_c \\ 0 & \text{si } |r| \geq r_c \end{cases} \quad (2-61)$$

Elle permet de contrôler la forme des pseudo-états et des projecteurs dans la région d'augmentation.

Le pseudo-hamiltonien $\tilde{H}(r)$ s'écrit :

$$\tilde{H}(r) = -\frac{1}{2} \nabla_r^2 + \tilde{V}_{eff}(r) \quad (2-62)$$

Et le pseudopotentiel effectif a pour expression :

$$\tilde{V}_{eff}(r) = \tilde{V}_{loc}(r) + \int \frac{\tilde{n}_v(r') + \hat{n}_v(r') + \tilde{n}_c(r')}{|r - r'|} d^3 r' + V_{ex}[\tilde{n}_v(r) + \tilde{n}_c(r)] \quad (2-63)$$

On remarque que, par rapport au potentiel effectif de l'équation (2-57), le pseudopotentiel effectif $\tilde{V}_{eff}(r)$ substitue le potentiel externe $V_{ext}(r)$ par un pseudopotentiel $\tilde{V}_{loc}(r)$ qui traduit les effets électrostatiques du noyau et des électrons de cœur fixes sur les pseudo-états de valence. $\tilde{n}_v(r)$ est la pseudo-densité de valence déduite des pseudo-états de valence et $\hat{n}_v(r)$ est un terme qui corrige la charge totale de l'atome dans le potentiel d'interaction coulombien classique de *Hartree*. $\tilde{n}_c(r)$ est la pseudo-densité de cœur construite comme une distribution de densité à symétrie sphérique à l'intérieur de la région d'augmentation et égale à la densité de valence au-delà de r_c qui a pour valeur d'après la forme proposée par *Holzwarth et coll* [35].

$$\tilde{n}_c(r) = \Gamma e^{(-\gamma r^2)} \quad \text{si } |r| < r_c \quad (2-64)$$

$$\tilde{n}_c(r) = n_c(r) \quad \text{si } |r| \geq r_c \quad (2-65)$$

Avec Γ et γ des paramètres ajustables pour modifier la forme de la fonction.

On calcule enfin les projecteurs correspondants avec :

$$|\tilde{P}_j\rangle = \left(-\frac{1}{2} \nabla_r^2 + \tilde{V}_{eff}(r) - \varepsilon_j \right) |\tilde{\varphi}_j\rangle \quad (2-66)$$

Est nul en dehors de la sphère d'augmentation. Pour un moment angulaire L_j donné, il existe à priori plusieurs ensembles de fonctions de base et de projecteurs correspondant à des états pris en compte dans la distribution de valence,

Blöchl démontre que les projecteurs doivent vérifier la propriété

$$\langle \tilde{P}_{j,n_{jk}} | \tilde{\varphi}_{j,n_{jk'}} \rangle = \delta_{n_{jk}, n_{jk'}} \quad (2-67)$$

Cette condition qui forme la quatrième étape consiste à orthogonaliser le $k^{\text{ième}}$ ensemble des fonctions de base et des projecteurs de chaque moment angulaire L_j par rapport aux ensembles de plus faibles énergies $n_{jk'} < n_{jk}$.

Afin d'estimer la qualité physique des pseudopotentiels, des tests élémentaires doivent être effectués sur l'ensemble des fonctions de base et des projecteurs, (i) l'amplitude de chaque pseudo-fonction $\tilde{\varphi}_j(r)$ doit être égale à l'amplitude de la fonction $\varphi_j(r)$ au-delà de r_c et les amplitudes des pseudo-fonctions $\tilde{\varphi}_j(r)$ et des projecteurs $\tilde{P}_j(r)$ doivent être du même ordre de grandeur, (ii) les pseudo-fonctions d'ondes $\tilde{\varphi}_j(r)$ doivent présenter le bon nombre de nœuds. (iii) les dérivées logarithmiques de la pseudo-fonction d'onde de valence $|\tilde{\psi}\rangle$ et de la fonction d'onde de valence $|\psi\rangle$ doivent correspondre pour chaque moment angulaire. Ces quantités sont calculées au niveau du rayon d'augmentation et représentées en fonction de l'énergie. Elles permettent de juger directement la qualité du ou des ensembles de fonctions de base et de projecteurs construits pour chaque moment angulaire. En première approximation il est conseillé de se contenter d'un seul ensemble de fonctions de base et de projecteurs si la correspondance des dérivées logarithmiques correspondants n'est pas satisfaisante, il peut être nécessaire de rajouter un état non lié pour une énergie de référence

choisie au voisinage de la zone de non correspondance des dérivées logarithmiques afin de résoudre ce problème.

2.9 Conclusion

Dans ce chapitre nous avons présenté la méthodologie pour l'étude théorique des systèmes périodiques. Cette présentation a été entamée par l'équation de Schrödinger, les différentes approximations utilisées et les bases choisies. Nous avons d'autre part présenté les notions fondamentales concernant la construction de pseudopotentiels pour introduire la méthode PAW de *Blöchl* qui permet de produire des pseudopotentiels.

Dans le chapitre suivant, nous présentons une nouvelle méthode d'analyse basée cette fois ci sur l'exploration et l'extraction de connaissances à partir de données, appelée communément « *Datamining* ».

Références

- [1] P. Hohenberg, W. Kohn, "Inhomogeneous Electron Gas" Phys. Rev. B 136, 864 (1964).
- [2] W. Kohn, L.J. Sham. "Self Consistent Equations Including Exchange and Correlation Effect" Phys. Rev. 140 (1965) A1133.
- [3] C. Frayret, « Application de la théorie de la fonctionnelle de la densité à la modélisation de la diffusion de l'ion oxygène dans des électrolytes solides modèles et des conducteurs mixtes » Université de Bordeaux 1, 19/10/2004.
- [4] M. Born, R. Oppenheimer, Ann. Phys. (Leipzig) 84 (1927) 457.
- [5] D.R. Hartree, "The Wave Mechanics of an Atom with a Non-Coulomb Central Fields". Proc. Cambridge Philos. Soc.24, (1928) 89-110.
- [6] D.R. Hartree: The calculation of Atomic Structures. John Wiley and Sons,1957.
- [7] V. Fock: Zeitschrift für Physik, 61 (1930) 126.
- [8] J.C. Slater. "A simplification of the Hartree-Fock method". Phys. Rev. 81 (1951) 385-390.
- [9] L.H. Thomas, Proc. Cambridge Philos. Soc. 23 (1927) 542.
- [10] E. Fermi, Z. Phys. 48 (1928) 73.
- [11] J.C. Slater, J.B. Mann, T.M. Wilson, J.H.Wood. « Nonintegral Occupation Numbers in Transition Atoms in Crystals ». phys. Rev. 184 (1969) 672-694.
- [12] D.J.Singh. Planewaves, Pseudopotentials and the LAPW Method. Kluwer Academic Publishers, Boston (USA), (1994).
- [13] L.J. Sham, W. Kohn, Phys. Rev. 145 561(1966).
- [14] P.A.M. Dirac, Proc. Camb. Phil. Soc. 26 (1930) 376.
- [15] D.M. Ceperley, B.J. Alder. "Ground State of the Electron Gas by a Stochastic Method" Phys. Rev. Lett 45 (1980) 566-569.
- [16] G. Ortiz et P. Ballone « Correlation energy, structure factor, radial distribution function, end momentum distribution of the spin-polarized uniform electron-gas. Phys. Rev. B 50(3), 1994.
- [17] L.Hedin, B.I. Lundqvist, Journal Phys. C4 (1971) 2064.
- [18] S.J. Vosko, L. Wilk, N. Nussair, Can. J. Phys. 58 (1980) 1200.
- [19] R.M.Martin, electronic structure: Basic theory and practical methods, Cambridge university press, 2004.
- [20] J.P. Perdew, Y.Wang, "accurate and simple analytic representation of the electron-gas correlation energy". Phys.Rev. B 45 (1992)13244-13249.
- [21] J.P.Perdew, K.Burke, M. Ernzerhof, Phys. Rev. Lett. 77 (1996)18.
- [22] Z.Wu, R. E. Cohen, Phys.Rev. B73 (2006) 235116.

-
-
- [23] E. Wimmer, Computational Materials Design: a perspective for atomistic approaches, J. Computer-Aided materials Design, Vol. 1 (1993) 215.
- [24] D.J. Chadi and L. Cohen, Phys. Rev. B 8 (1973) 5747.
- [25] H.J. Monkhost, J.D. Pack : Special points for Brillouin-zone integrations, Phys. Rev. B 13(12) (1976) 5188.
- [26] J.C. Slater, wave Functions in a periodic Potential, Phys. Rev. 51 (1937) 846.
- [27] S.Cottenier, Density Functional Theory and the family of (L)APW-methods: a step-by-step in introduction (instituut voor Kern en Stralingsfysica, K.U.Leuven, Belgium) 2002.
- [28] D.Sing, L. Nordström Planewaves, pseudopotentials and LAPW-method, Kluwer Academic Publishing (1994), ISBN 0-7923-9421-7.
- [29] W.E. Pickett, Comput. Phys. Rep. 9 115 (1989).
- [30] D.R. Hammam, M. Schluter, C. Chiang, Phys. Rev. Lett.43 1494 (1979).
- [31] G.P. kerker « Non-singular atomic pseudopotentials for solid state application » J. Phys.C: solid St. Phys.13, L189 (1980).
- [32] D.Vanderbilt « soft self-consistent pseudopotentials in a generalized eigenvalue formalism » Phys. Rev. B41, 7892 (1990).
- [33] P.E.Blöchl «Projector augmented wave method » Phys. Rev. B50, 17953 (1994).
- [34] F. Lallet, « Etude numériques et expérimentales de systèmes auto-organisés à différentes échelles : exemple d'îlots YSZ nanométriques et de particules de diamant micrométriques» Université de Limoges, 02/06/2008.
- [35] N.A.W.Holzwarth, G.E. Matthews, R.B.Dunning, A.R. Tackett and Y. Zeng. «comparison of the projector augmented-wave, pseudopotential, and linearized augmented-plane-wave formalisms for density-functional calculations of solids»

CHAPITRE III

DATAMINING

DATAMINING

3.1 Introduction

L'analyse de données permet d'obtenir des informations synthétiques à partir de données recueillies. Il existe de nombreuses problématiques et autant de méthodes permettant ou essayant d'y répondre. Par ailleurs, les moyens informatiques modernes ont permis de produire et d'archiver d'énormes masses de données numériques au moins à partir du début des années 1990. Si ces données sont généralement collectées pour rendre un service donné ou répondre à une question précise, la question de l'analyse de grands volumes de données s'est imposée. L'essor d'une nouvelle discipline scientifique appelée « extraction de connaissances à partir de données (ECD) » (*Knowledge Discovery in Databases (KDD)*) [1] qui est un domaine de recherche tentant de répondre à cette question en mettant au point des outils et méthodes de fouille de données (*outils de DataMining*) capables d'analyser les grands volumes de données. Toutefois ces méthodes d'analyse sont issues du domaine de la statistique traditionnelle : histogramme, courbe de concentration, régression linéaire, analyse des corrélations, classification ascendante hiérarchique. D'un autre côté, certaines de ces méthodes ont été adaptées pour le traitement et l'analyse de données et rentrent dans le cadre de l'exploitation et d'analyse et font partie intégrante du Datamining, tels que l'Analyse en Composante Principale (ACP), les réseaux de neurones (*Artificial Neural Network ; ANN*) et les algorithmes génétiques (*genetic algoritm ; GA*) [2,3].

Dans ce chapitre, on s'intéresse uniquement à deux techniques de l'exploitation et l'analyse de données, qui sont les algorithmes génétiques (AG) et les réseaux de neurones artificiels (RNA). Ces deux méthodes ont connu ces dernières années un développement considérable grâce à l'augmentation vertigineuse de la puissance des calculateurs et notamment suite à l'apparition des architectures massivement parallèles qui exploitent leur parallélisme intrinsèque.

Ce chapitre est organisé en trois parties, la première partie est consacrée à la définition du Datamining, et ses différentes techniques. La deuxième partie fait l'objet d'une étude détaillée sur l'emploi des réseaux de neurones. Après un bref historique et une représentation de quelques notions générales sur les réseaux de neurones, nous présentons les deux grandes familles de structures neuronales les plus utilisées. Ensuite nous nous intéressons plus

particulièrement au perceptron multicouche (*Multilayer Perceptron* : MLP). Nous abordons également le problème d'apprentissage des paramètres de ces structures, ainsi que le traitement des données. Et enfin la troisième et la dernière partie est consacrée aux algorithmes génétiques comme étant une méthode d'optimisation. Après une représentation générale du principe de fonctionnement d'un algorithme génétique, nous ferons une description détaillée de ses différentes composantes.

3.2 Datamining

Le Datamining est l'extraction des connaissances et d'informations d'une large base de données [4]. Aussi connue sous le nom « fouille de données » [5].

Il a pour vocation d'être utilisé dans un environnement professionnel, en vue de trouver des structures originales et des corrélations informelles entre les données, il permet de mieux comprendre les liens entre des phénomènes en apparence distinctes et d'anticiper des tendances encore peu discernables restituant de façon concise l'essentiel de l'information utile pour aider à la prise de décision.

Considérée comme l'une des sciences les plus révolutionnaires de la prochaine décennie, elle est en plein essor, cette technologie est motivée par de nouvelles techniques permettant d'exploiter, d'analyser, de comprendre et de traiter de grandes quantités de données [6]. Utilisée dans plusieurs domaines [1] tels que la gestion de grandes entreprises, analyse de la consommation, sciences de l'ingénierie, la santé, la socio démographie, science sociale ou l'éducation, et les finances comme par exemple pour mieux connaître et prédire le comportement des clients (suivre l'évolution des prêts au niveau des clients potentiels et de leurs remboursement) [7-11]. Récemment, le Datamining trouve ses applications dans un large éventail de sujets y compris dans le domaine des sciences et technologies [5], plusieurs travaux ont été publiés concernant ses applications.

Cependant le processus d'extraction de connaissances et d'informations suit certaines étapes principales (voir figure 3-1). (i) Formaliser un problème que le *scientifique* cherche à résoudre, (ii) l'accès aux données, (iii) préparation des données en vue des traitements et utilisation, (iv) modélisation des données en leur appliquant des algorithmes d'analyse, et enfin l'évaluation et validation des connaissances ainsi extraites des analyses [5].

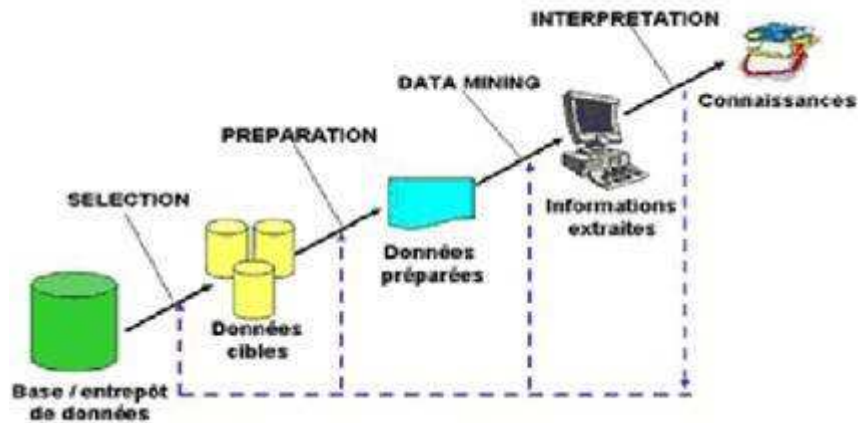


Figure 3-1: Schéma du processus du Datamining.

3.3 Techniques et méthodes

L'exploitation des données se propose d'utiliser un ensemble de techniques ou méthodes issues de disciplines scientifiques diverses tels que l'intelligence artificielle (AI) ou l'informatique. Pour construire des modèles à partir des données, il existe deux grandes familles de, les méthodes descriptives et les méthodes prédictives.

3.3.1 Méthodes non supervisées (techniques descriptives)

Ces méthodes dites aussi *techniques descriptives* permettent de travailler sur un ensemble de données dans lequel aucune variable explicative n'a d'importance particulière par rapport aux d'autres, c'est-à-dire dans un ensemble de données aucune variable n'est considérée comme cible, ces techniques sont utilisées pour dégager d'un ensemble de variables des groupes homogènes et de réaliser une compression d'informations. Parmi ces techniques descriptives la plus utilisée est l'analyse factorielle illustrée sur la figure 3-2, l'idée générale est de dégager des variables cachées appelées « facteurs » de l'ensemble de mesures ou données. Dans cette analyse on part du principe que si les données sont dépendantes entre elles c'est parce qu'elles sont liées à des facteurs qui leurs sont communs, l'intérêt des facteurs réside dans le fait qu'un nombre réduit de facteurs explique aussi bien les données que l'ensemble des variables. Ces analyses factorielles se décomposent principalement en ; Analyse en Composantes Principales (ACP), Analyse en Composantes Indépendantes et Analyse en Correspondance Multiples.

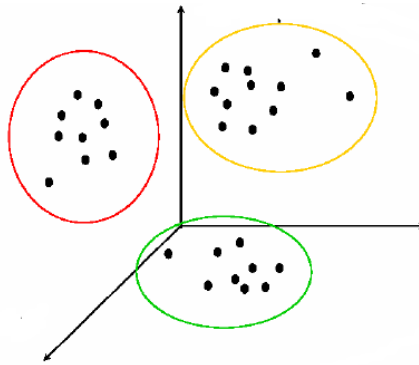


Figure 3-2 : Exemple de techniques non supervisées.
(Analyse en Composantes Principale)

3.3.2 Méthodes supervisées (techniques prédictives)

Ces méthodes dites aussi *techniques prédictives* (figure 3-3) servent à expliquer ou à prévoir un ou plusieurs phénomènes observables, leurs objectifs est de s'intéresser à une ou plusieurs variables de la base de données définies comme étant les cibles de l'analyse. Les principales techniques supervisées existantes sont :

- Arbres de décision
- Réseaux de neurones
- Régressions linéaires et non linéaires
- Les algorithmes génétiques
- Réseau Bayésien

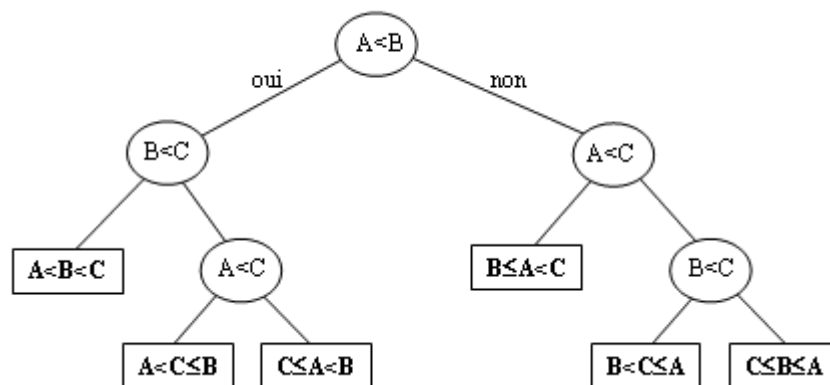


Figure 3-3: Exemple de Techniques supervisées.
(Arbre de décision : algorithme de tri par insertion de trois éléments)

3.4 Réseaux de neurones artificiels

3.4.1 Historique

La première modélisation du réseau biologique a été faite en 1943 par *W. McCulloch* et *W. Pitts*, ils ont montré que les réseaux de neurones simples peuvent réaliser des fonctions logiques, arithmétiques et symboliques complexes (tout en moins au niveau théorique). En 1949, *D. Hebb*, physiologiste américain présente dans son ouvrage « *The Organization of Behavior* » une règle d'apprentissage, que de nombreux modèles de réseaux aujourd'hui s'inspirent encore de la règle de *Hebb*. En 1958, les premiers succès ont vu le jour par les travaux de *F. Rosenblatt* qui a développé le modèle du Perceptron. C'est un réseau de neurones inspiré du système visuel basé sur la reconnaissance de formes. Il possède deux couches de neurones : une couche de perception et une couche liée à la prise de décision. C'est le premier système artificiel capable d'apprendre par expérience. Notons qu'à cet époque les moyens informatiques sont limités et une prouesse technologique de réussir à faire fonctionner cette machine que de quelques minutes. Dans la même période, le modèle l'Adaline (ADaptive LINar Element) a été présenté par *B. Widrow* et *Hoff*, et qui sera par la suite un modèle de base des réseaux multicouches. En 1969 *M. Minsky* et *S. Papert* publient une critique sur la limitation théorique du Perceptron, notamment concernant l'impossibilité de traiter par ce modèle des problèmes non linéaires. Cela va avoir une incidence sur la recherche dans ce domaine et un abandon financier. Après des années d'ombre entre 1967 et 1982, un nouveau rebondissement a vu le jour grâce aux travaux du physicien *J. J. Hopfield* à qui on doit un nouveau intérêt pour les réseaux de neurones artificiels, et la théorie des réseaux de neurones devient respectable, et qui n'est plus l'apanage d'un certain nombre de psychologues et neurobiologistes. Dès lors, au début des années 80 les premiers modèles apparaissent avec la rétro propagation de gradient, qui est un algorithme d'apprentissage adapté aux réseaux de neurones multicouches (aussi appelés Perceptron multicouches). Ils traitent de manière satisfaisante les problèmes de limitation pour les perceptrons, dès cette découverte, nous avons la possibilité de réaliser une fonction non linéaire d'entrée/sortie sur un réseau en décomposant cette fonction en une suite d'étapes linéairement séparables. De nos jours, les réseaux multicouches et la rétro propagation de gradient restent le modèle le plus étudié et le plus productif au niveau des applications.

3.4.2 Neurone formel

Un neurone formel (ou neurone) comme le montre la figure 3-4, est un opérateur mathématique qui transforme une ou plusieurs variables d'entrées (x) en une ou plusieurs valeurs de sorties (y). Basé sur la structure et le fonctionnement d'un neurone biologique [12], il réalise une fonction algébrique, non linéaire, de ses variables d'entrées. Dans une version simple, un neurone calcule la somme pondérée des entrées reçues, puis applique à cette valeur une fonction d'activation, généralement non linéaire, la valeur finale obtenue est la sortie du neurone.

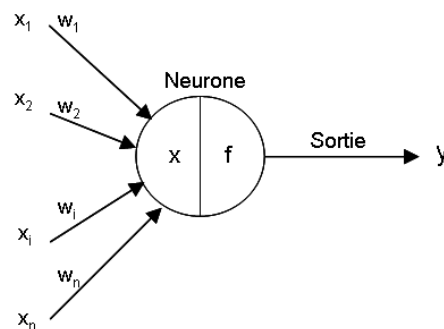


Figure 3-4 : Un neurone formel.

Les entrées du neurone sont désignées par x_i ($i = 1, \dots, n$), les paramètres w_i sont appelés poids synaptiques ou tout simplement poids, la sortie du neurone est donnée par :

$$\text{Sortie} = f(x) \text{ avec } x = \sum_{i=1}^n w_i x_i \quad (3-1)$$

Où x est l'état d'activation du neurone (également appelé état ou activité), f est la fonction d'activation du neurone, elles sont généralement croissantes et bornées. Les plus connus sont la fonction signe, la fonction linéaire saturée et la fonction sigmoïde.

Il existe plusieurs types de fonctions d'activations possibles (voir figure 3-5). Toutes ces fonctions donnent généralement de meilleurs résultats. Il est toutefois possible, dans certains cas particuliers d'utiliser des fonctions plus spécifiques liées à leur contexte d'utilisation afin d'obtenir des meilleurs résultats. Le choix de celles-ci se fait en fonction de la position du neurone dans le réseau et aussi en fonction des domaines d'applications [13].

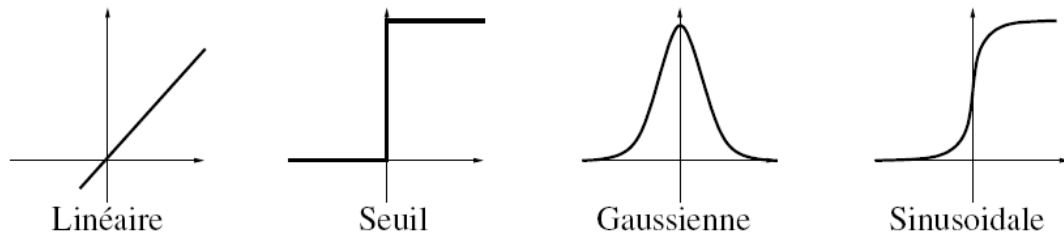


Figure 3-5 : Différentes fonctions d'activation usuelles

3.4.3 Réseaux de neurones artificiels (RNA)

Un réseau de neurones artificiels (RNA) (*Artificial Neural Networks* : ANN) est un système de neurones artificiels interconnectés, il permet d'émuler le fonctionnement de base d'un réseau de neurones biologiques dans le but de modéliser un système particulier [4,14]. Ils sont connus comme des outils puissants pour la modélisation, en particulier lorsque les relations entre les données sont méconnues. Ils sont généralement acceptés comme une technologie offrant une autre façon de simuler des problèmes ambigus et complexes [15]. Parmi les différents réseaux de neurones existants, on peut distinguer deux grandes familles : les réseaux de neurones non bouclés et les réseaux de neurones bouclés.

✚ Les réseaux non bouclés.

Dans lesquels l'information circule des entrées vers les sorties sans bouclage (voir figure 3-6), se sont des systèmes statiques, la sortie du réseau est donc une fonction de variables d'entrées pour laquelle le temps n'a pas de rôle fonctionnel. Utilisés principalement pour effectuer des tâches de classification, ou de modélisation statique de processus, dont la relation réalisée entre l'entrées-sorties est une fonction algébrique [16].

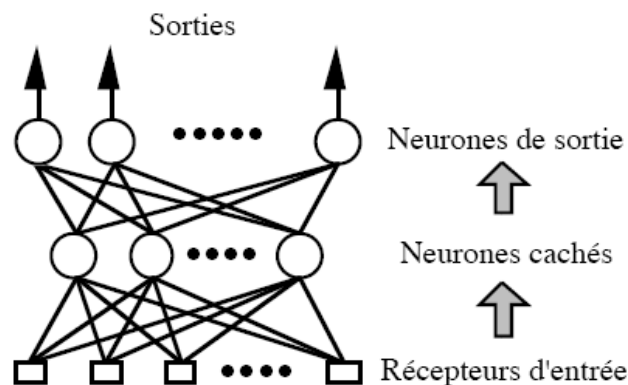


Figure 3-6 : Réseau de neurones non bouclé.

✚ Les réseaux bouclés

Comme les réseaux non bouclés, la sortie du réseau est une fonction de n variables externes dites variables exogènes (voir Figure 3-7), elle est également fonction d'une ou plusieurs valeurs passées des sorties du même réseau de neurones qui sont appelées variables d'état. Ainsi, le temps joue un rôle fonctionnel puisque la sortie du réseau dépend d'une ou plusieurs variables de ses sorties passées. On a donc bien un système dynamique. Utilisés comme filtres non linéaires pour la modélisation et la commande de processus, et la relation réalisée entre l'entrées-sorties est un ensemble d'équations aux différences couplées [17].

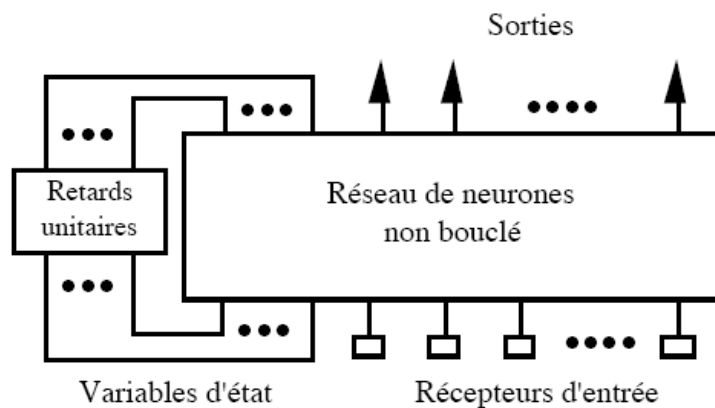


Figure 3-7 : Réseau de neurones bouclé.

3.4.4 Perceptrons multicouches

Il existe plusieurs types de réseaux de neurones, tel que le perceptron multicouche (MLP), le réseau de neurones probabiliste (probabilistic neural network), réseau de neurones à régression généralisée (generalized regression neural network) et le Kohomen. Parmi ces différents types, le perceptron multicouches reste le plus populaire et le plus utilisé [18,19], développé en 1940 par *Mc Culloch* et *Pitts* [19], il est constitué d'une couche d'entrée formée de nœuds, une ou plusieurs couches cachées et une couche de sortie (voir figure 3-6). Le nombre de nœuds dans la couche d'entrée et de sortie dépend respectivement du nombre de variables d'entrées et de sorties [20].

Le MLP est très utilisé en identification et en contrôle, il constitue un approximateur universel. De récentes recherches montrent qu'il peut être entraîné de manière à approximer n'importe quelle fonction entrées-sorties sous réserve de mettre suffisamment de neurones dans la couche cachée et d'utiliser des sigmoïdes pour les fonctions d'activation [21].

3.4.5 Traitement des données

Pour qu'un réseau de neurones artificiel effectue une tâche donnée, un traitement préalable des variables d'entrées est indispensable afin de ramener toutes les variables à des ordres de grandeurs similaires. Ainsi les variables ayant des valeurs faibles pourront être prises en compte au même titre que les variables ayant des valeurs fortes. De plus, le traitement des données permet d'éviter d'avoir de fortes valeurs pouvant entraîner la saturation de la fonction d'activation, surtout s'il s'agit d'une fonction sigmoïde. Donc les valeurs doivent être normalisées de façon à ne pas atteindre les minimums et les maximums de la fonction d'activation [15].

3.4.6 Apprentissage des réseaux de neurones

L'apprentissage dans le contexte des réseaux de neurones est un processus de modification des poids de connexion de telle manière que les sorties du réseau de neurones soient aussi proches que possible des sorties 'désirées' (voir figure 3-8) [22,23], ceci dans le but d'adapter le traitement effectué par le réseau de neurones pour une tâche particulière. On distingue trois familles d'apprentissage :

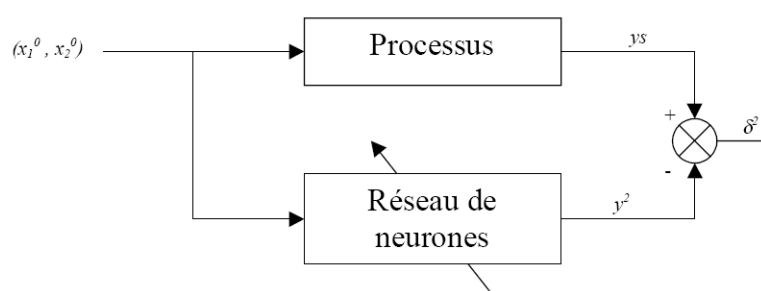


Figure 3-8 : Schéma synoptique du procédé d'apprentissage du réseau de neurone.

🚦 Apprentissage supervisé

Dans ce mode d'apprentissage, un superviseur (expert humain) fournit une valeur ou un vecteur de sortie appelé cible ou sortie désirée que notre réseau de neurones doit associer à un vecteur d'entrée. Ce mode d'apprentissage consiste dans ce cas à ajuster les paramètres du réseau afin de minimiser l'erreur entre la sortie désirée et la sortie réelle du réseau.

🚦 Apprentissage non supervisé

Dans cet apprentissage, les données ne contiennent pas d'informations sur la sortie désirée. Il n'y a pas de superviseur. La tâche du réseau consiste par exemple à créer des regroupements de données selon des propriétés communes (classification).

Apprentissage hybride

Ce mode d'apprentissage, est plus rare et est encore mal ou peu exploré. Cette approche consiste à combiner une ou plusieurs techniques d'optimisations. Comme par exemple une hybridation combinant l'algorithme génétique et le réseau de neurones. (*Le réseau de neurone aura la qualité de fonction d'évaluation*). Ce mode d'apprentissage sera développé dans le chapitre V.

3.4.7 Rétropropagation du gradient

La rétropropagation est une méthode basée sur un algorithme spécifique aux réseaux de neurones qui consiste à ajuster les poids w_i en fonction du gradient de la fonction de coût [24]. Elle a été développée en particulier par *Rumelhart* et *McClelland* [25]. Cet algorithme repose sur la minimisation de l'erreur quadratique entre les sorties calculées et celles souhaitées. L'objectif est de modifier les poids du réseau dans le sens contraire du gradient du critère de performance. Dans ce qui suit, nous allons présenter les équations constituant l'algorithme en utilisant un réseau multicouche.

Considérons le modèle du neurone j présenté sur la figure 3-9, et pour alléger l'exposé, on suppose que l'apprentissage se fait à chaque présentation d'un couple entrée/sortie de l'ensemble d'apprentissage.

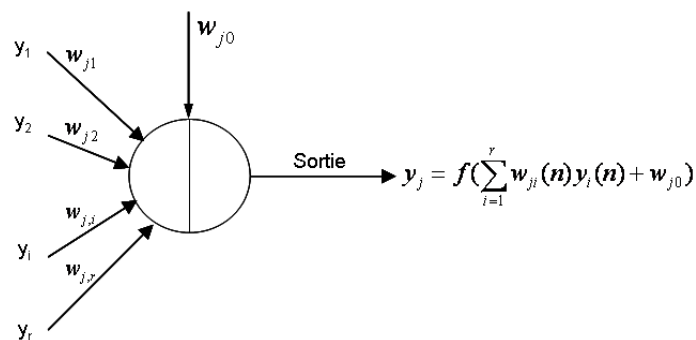


Figure 3-9 : Modèle du neurone j

Cas de la couche de sortie

Soit l'erreur observée $e_j(n)$ pour le neurone de sortie j et la donnée d'entraînement n :

$$e_j(n) = d_j(n) - y_j(n) \quad (3-2)$$

$d_j(n)$ correspond à la sortie désirée du neurone j et $y_j(n)$ à la sortie observée.

Par conséquent l'erreur totale observée sur l'ensemble des neurones de sortie :

$$E(n) = \frac{1}{2} \sum e_j^2(n) \quad (3-3)$$

La sortie $y_j(n)$ du neurone j est définie par :

$$y_j(n) = f(v_j(n)) = f\left(\sum_{i=1}^r w_{ji}(n) y_i(n) + w_{j0}\right) \quad (3-4)$$

Où f est la fonction d'activation du neurone, $v_j(n)$ est la somme pondérée des entrées du neurone j , $w_{ji}(n)$ le poids du lien connectant le neurone j au neurone i de la couche précédente, w_{j0} le biais du neurone j , et $y_i(n)$ la sortie du neurone i de la couche précédente. On suppose que la couche précédente contient r neurones

La correction des poids par descente du gradient de l'erreur quadratique moyen s'écrit :

$$\Delta w_{j,i} = -\eta \frac{\partial E(n)}{\partial w_{j,i}} \quad (3-5)$$

Avec $0 \leq \eta \leq 1$ représentant un taux d'apprentissage (ou gain de l'algorithme)

Sachant que l'erreur du neurone j dépend des neurones de la couche précédente, en utilisant

la règle du chaînage des dérivées partielles $\left(\frac{\partial f}{\partial x} = \frac{\partial f}{\partial y} \cdot \frac{\partial y}{\partial x}\right)$

$$\frac{\partial E(n)}{\partial w_{ji}(n)} = \frac{\partial E(n)}{\partial e_j(n)} \frac{\partial e_j(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \frac{\partial v_j(n)}{\partial w_{ji}(n)} \quad (3-6)$$

Et

$$\frac{\partial E(n)}{\partial w_{j0}(n)} = \frac{\partial E(n)}{\partial e_j(n)} \frac{\partial e_j(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \frac{\partial v_j(n)}{\partial w_{j0}(n)} \quad (3-7)$$

Evaluons chacun des termes du gradient.

$$\frac{\partial E(n)}{\partial e_j(n)} = \frac{\partial \left[\frac{1}{2} \sum_k e_k^2(n) \right]}{\partial e_j(n)} \quad (3-8)$$

$$= \frac{1}{2} \frac{\partial e_j^2(n)}{\partial e_j(n)} \quad (3-9)$$

$$= e_j(n) \quad (3-10)$$

$$\frac{\partial e_j(n)}{\partial y_j(n)} = \frac{\partial [d_j(n) - y_j(n)]}{\partial y_j(n)} \quad (3-11)$$

$$= -1$$

$$\frac{\partial y_j(n)}{\partial v_j(n)} = \frac{\partial \left[\frac{1}{1 + e^{-v_j(n)}} \right]}{\partial v_j} \quad (3-12)$$

$$= \frac{e^{-v_j(n)}}{(1 + e^{-v_j(n)})^2} \quad (3-13)$$

$$= y_j(n) \left[\frac{e^{-v_j(n)}}{1 + e^{-v_j(n)}} \right] \quad (3-14)$$

$$= y_j(n) \left[\frac{e^{-v_j(n)} + 1}{1 + e^{-v_j(n)}} - \frac{1}{1 + e^{-v_j(n)}} \right] \quad (3-15)$$

$$= y_j(n) [1 - y_j(n)] \quad (3-16)$$

Et finalement :

$$\frac{\partial v_j(n)}{\partial w_{j,i}(n)} = \frac{\partial \left[\sum_{l=0}^r w_{j,l}(n) y_l(n) \right]}{\partial w_{j,i}(n)} \quad (3-17)$$

$$= \frac{\partial [w_{j,i}(n) y_i(n)]}{\partial w_{j,i}(n)} \quad (3-18)$$

$$= y_i(n) \quad (3-19)$$

Et

$$\frac{\partial v_j(n)}{\partial w_{j,0}(n)} = \frac{\partial \left[\sum_{l=0}^r w_{j,l}(n) y_l(n) + w_{j,0} \right]}{\partial w_{j,0}(n)} \quad (3-20)$$

$$= 1$$

Nous obtenons donc :

$$\frac{\partial E(n)}{\partial w_{j,i}(n)} = -e_j(n) y_j(n) [1 - y_j(n)] y_i(n) \quad (3-21)$$

L'apprentissage des poids de la couche de sortie :

$$\Delta w_{j,i} = -\eta \frac{\partial E(n)}{\partial w_{j,i}} = \eta \delta_j(n) y_i(n) \quad (3-22)$$

$$\text{Avec } \delta_j(n) = e_j(n) y_j(n) [1 - y_j(n)] \quad (3.23)$$

L'apprentissage des biais de la couche de sortie :

$$\Delta w_{j,0} = -\eta \frac{\partial E(n)}{\partial w_{j,0}} = \eta \delta_j(n) \quad (3-24)$$

✚ Cas d'une couche cachée

Considérons maintenant le cas des neurones sur la dernière couche cachée, n désignera la donnée d'entraînement c'est-à-dire un couple de vecteurs d'entrées et de sorties désirées, l'objectif sera toujours d'adapter les poids de la couche courante en minimisant la somme des erreurs sur les neurones de la couche de sortie. Les indices i et j désigneront respectivement un neurone sur la couche précédente et un neurone sur la couche courante, l'indice k servira maintenant à désigner un neurone sur la couche suivante.

Reprenons l'expression (3-6) de la dérivée partielle de l'erreur quadratique $E(n)$ par rapport à w_{ji} . Les deux derniers termes de cette équation restent inchangés, par rapport aux résultats obtenus pour la couche de sortie, seul le premier terme requiert d'être évalué :

$$\frac{\partial E(n)}{\partial y_j(n)} = \frac{\partial \left[\frac{1}{2} \sum_k e_k^2(n) \right]}{\partial y_j(n)} \quad (3-25)$$

$$= \frac{\partial}{\partial y_j(n)} \frac{1}{2} \sum_k e_k^2(n) = \sum_k e_k(n) \frac{\partial e_k(n)}{\partial y_j(n)} \quad (3-26)$$

$$= \sum_k \left[e_k(n) \frac{\partial e_k(n)}{\partial v_k(n)} \frac{\partial v_k(n)}{\partial y_j(n)} \right] \quad (3-27)$$

$$= \sum_k \left[e_k(n) \frac{\partial [d_k(n) - f(v_k(n))]}{\partial v_k(n)} \frac{\partial \left[\sum_l w_{kl}(n) y_l(n) \right]}{\partial y_j(n)} \right] \quad (3-28)$$

$$= \sum_k [e_k(n)(-y_k(n)[1 - y_k(n)])w_{kj}] \quad (3-29)$$

$$\frac{\partial E(n)}{\partial y_j(n)} = -\sum_k \delta_k(n) w_{kj}(n) \quad (3-30)$$

En substituant l'équation 3-30 dans l'équation 3-6 on obtient :

$$\frac{\partial E(n)}{\partial w_{ji}(n)} = -y_j(n)[1 - y_j(n)] \left[\sum_k \delta_k(n) w_{kj}(n) \right] y_i(n) \quad (3-31)$$

Et finalement l'apprentissage de la couche cachée s'écrit :

$$\Delta w_{j,i} = -\eta \frac{\partial E(n)}{\partial w_{j,i}} = \eta \delta_j(n) y_i(n) \quad (3-32)$$

$$\text{Avec } \delta_j(n) = y_j(n)[1 - y_j(n)] \sum_k \delta_k(n) w_{kj}(n) \quad (3-33)$$

Et l'apprentissage pour le biais des couches cachées s'écrit :

$$\Delta w_{j,0} = -\eta \frac{\partial E(n)}{\partial w_{j,0}} = \eta \delta_j(n) \quad (3-34)$$

Ces équations sont valides pour toutes les couches cachées.

Dans cette partie nous avons décrit, avec un certain détail la méthode de rétropropagation pour les couches de sorties et les couches cachées et déterminé l'erreur entre la sortie désirée et la sortie calculée avec le réseau de neurones. On réitère cette d'opération sur un ensemble d'apprentissage tant que l'erreur de sortie reste supérieure à un certain seuil, et tant que l'apprentissage reste efficace et ne rentre pas dans la phase de sur-apprentissage qui rend le réseau de neurones incapable de remplir correctement son rôle d'interpolateur.

3.5 Les algorithmes génétiques

Les phénomènes physiques ou biologiques ont été la source d'inspiration de nombreux algorithmes de calcul numérique pour la modélisation et particulièrement pour l'optimisation. Les plus connus sont les algorithmes génétiques (AG), basés sur l'évolution biologique et de la génétique [23], ils tentent de simuler le processus d'évolution naturelle suivant le modèle darwinien [26] dans un environnement donné. Ils utilisent un vocabulaire similaire à celui de la génétique naturelle.

Par analogie avec l'évolution naturelle, les AG font évoluer un ensemble de solutions candidates, appelé « population d'individus ». Un « individu » n'est autre qu'une solution possible du problème à résoudre. Chaque individu de cette population se voit attribuer une fonction appelée fonction d'adaptation « *fitness* » qui permet de mesurer sa qualité ou son poids, cette fonction d'adaptation peut représenter la fonction objectif à optimiser. Ensuite, une fois les meilleurs individus de cette population sont sélectionnés, ils subissent des croisements et des mutations, et une nouvelle population de solutions est produite par la génération suivante. Ce processus se poursuit génération après génération, jusqu'à ce que le critère d'arrêt soit atteint, comme par exemple le nombre maximal de générations.

Les algorithmes génétiques sont des méthodes d'optimisations stochastiques maintenant bien connues, ils diffèrent de la plupart des techniques d'optimisation en raison de leur recherche globale d'une population de solutions plutôt que d'une seule solution [27]. Les premiers travaux ont été initialement développés par *John Holland* (1975) [28] en développant les principes fondamentaux des algorithmes génétiques dans le cadre de

l'optimisation mathématique. La parution en 1989 de l'ouvrage de référence de *D.E.Goldberg* [29] intitulé « Genetic Algorithms in Search optimization and Machin learning », qui décrit l'utilisation de ces algorithmes dans le cadre de résolution de problème concret, a permis de mieux faire connaître les AG et de les introduire dans la communauté scientifique, il a marqué le début d'un nouvel intérêt pour cette technique d'optimisation.

De telles méthodes peuvent permettre en calcul scientifique comme dans bien d'autres domaines (économie, finances, planification,...etc.) de dépasser certaines limitations des méthodes classiques. Elles sont en effet capables de travailler sur des espaces discrets aussi bien que continus.

3.5.1 Principe d'un algorithme génétique

Un algorithme génétique débute par une population initiale de taille déterminée, tous solutions de la fonction à optimiser [30] convenablement, marquées par un codage qui les identifie complètement. Une procédure d'évaluation est appliquée afin de déterminer la force de chaque individu de la population. Vient ensuite une phase de sélection aléatoire selon le principe « des lois de survie » selon lesquelles seuls les individus les mieux adaptés auront une longévité suffisante pour générer une descendance. Les individus peu adaptés auront une tendance à disparaître. Ensuite une phase de recombinaison les leurs appliquée (opérateurs de croisement et de mutation) qui génèrent une nouvelle population d'individus, qui ont de bonnes chances d'être plus forts que ceux de la génération précédente. De génération en génération, la force des individus de la population augmente et après un certain nombre d'itération, la population est entièrement formée d'individus tous forts, soit de solutions quasi-optimales du problème posé. La figure 3-10 présente le principe d'un AG standard.

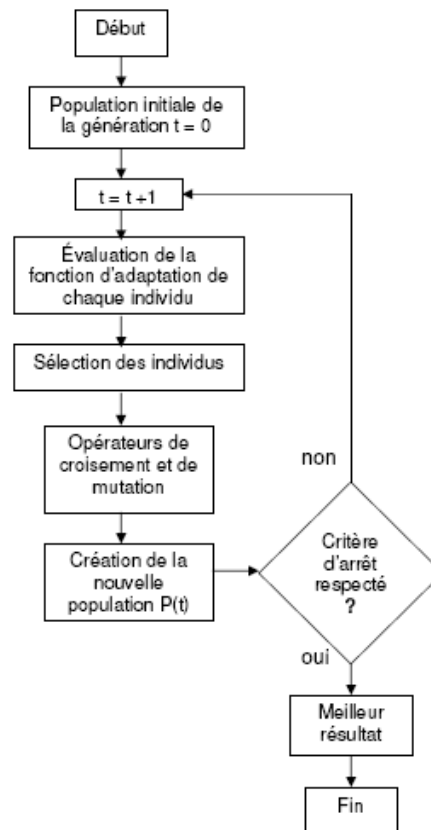


Figure 3-10 : Principe général des algorithmes génétiques standards.

3.5.2. Représentation des individus et codage

Premièrement, il faut représenter les différents états possibles de la variable dont on cherche la valeur optimale sous forme utilisable pour un algorithme génétique : c'est le codage. Cela permet d'établir une connexion entre la valeur de la variable et les individus de la population, de manière à imiter la transcription génotype - phénotype qui existe dans le monde vivant. Il existe principalement trois types de codage : le codage binaire, le codage réel et le codage Gray.

3.5.2.1 Le codage binaire

Ce type de codage a été largement utilisé dans le domaine des AG, néanmoins il présente quelques inconvénients pour des AG à problèmes d'optimisation de grande dimension et à haute précision numérique. Ces AG basées sur des chaînes binaires ont de faibles performances comme le montre *Michalewicz* (1992), et la distance de *Hamming* entre deux nombres voisins peut être assez grande dans le codage binaire. Par exemple l'entier 7 correspond à la chaîne 0111 et la chaîne 1111 correspond à l'entier 8, or la distance de

Hamming entre ces deux chaînes est de 4, ce qui crée bien souvent une convergence, et non pas l'obtention de la valeur optimale.

3.5.2.2 Le codage réel

Il a le mérite d'être simple, chaque chromosome est en fait un vecteur dont les composantes sont des paramètres du processus d'optimisation. Par exemple, si on cherche l'optimum d'une fonction de n variables $f(x_1, x_2, \dots, x_n)$, on peut utiliser tout simplement un chromosome ch contenant les n variables (figure 3-11), avec ce type de codage la procédure d'évaluation des chromosomes est plus rapide vu l'absence de l'étape de transcodage (du binaire vers le réel). Les résultats donnés par *Michalewicz* montrent que la représentation réelle aboutit souvent à une meilleure précision et un gain important en terme de temps d'exécution.

$$ch: \begin{array}{|c|c|c|c|c|} \hline x_1 & x_2 & \cdots & x_{n-1} & x_n \\ \hline \end{array}$$

Figure 3-11: exemple de codage réel.

3.5.2.3 Le codage de Gray

Egalement appelé « binaire réfléchi » est un type de codage binaire permettant de ne modifier qu'un seul bit à la fois quand un nombre est augmenté d'une unité. Utilisé surtout lors de la conception des circuits logiques. Le nom du code vient de l'ingénieur *Franck Gray* qui a déposé un brevet sur ce code en 1953.

3.5.2.4 Initialisation

L'initialisation sert à constituer la population initiale. C'est une étape importante car si la population n'est pas uniformément répartie au départ sur tout l'espace de recherche, l'évolution risque à se concentrer sur un optimum local duquel il peut être difficile de sortir. On essaye de créer une population initiale aussi diversifiée que possible, cette diversité de la population doit être entretenue aux cours des générations afin d'explorer le plus largement possible l'espace de recherche par des opérateurs génétiques qui simulent les processus naturels appliqués par la nature aux espèces vivantes. Ces opérateurs génétiques sont la sélection, le croisement et la mutation.

3.5.2.5 Fonction d'évaluation

Un algorithme génétique nécessite généralement la définition d'une fonction rendant compte de la pertinence des solutions potentielles, à chaque solution on associe une fonction performance appelée « *fitness* » reliée à la valeur de la fonction objectif. Cette fonction performance décrit le mérite de l'individu qui est représenté par un chromosome.

La fonction performance est très importante pour un algorithme génétique. En effet, pour que les algorithmes génétiques se comportent bien, nous devons trouver une manière de formuler des fonctions performance ne comportant pas trop de maxima locaux et ne présentant pas de maximum local isolé. Cette construction de fonctions performance est évidente pour certains problèmes. Pour les problèmes de maximisation par exemple, la fonction mérite peut être égale à la fonction objectif, par contre pour les problèmes de minimisation l'objectif est de trouver des solutions pour lesquelles la fonction objectif atteint des valeurs minimales. Dans ce cas, la fonction performance choisie est la réciproque de la fonction objectif. Dans tous les cas, l'algorithme génétique cherche à maximiser la fonction performance qui dans le cadre d'un problème de minimisation, prend la forme suivante :

$$Fitness(ch_i) = \frac{1}{f(ch_i)} \quad (3-35)$$

Où $f(ch_i)$ représente la fonction objectif évaluée pour l'individu ch_i .

3.5.2.6 Sharing

Le sharing consiste à ajuster la *fitness* des individus pour éviter qu'ils se concentrent dans une niche principale (optimum global). La technique de partage de la fonction *fitness* (*fitness sharing*), introduite par *Goldberg* et *Richardson* (*Goldberg* et *Richardson*, 1987), réduit la *fitness* de chaque individu d'un facteur correspondant environ au taux d'agrégation de la population autour de son voisinage :

$$f'(ch_i) = \frac{f(ch_i)}{m_i} \quad (3-36)$$

Où m_i : compteur de niche qui se calcule de la manière suivante :

$$m_i = \sum_{j=1}^N sh(d_{ij}) \quad (3-37)$$

Où N désigne la taille de la population et sh mesure la similarité entre deux individus i et j en fonction de la distance d_{ij} et le rayon de niche σ_{shar} :

$$sh(d_{ij}) = \begin{cases} 1 - \left(\frac{d_{ij}}{\sigma_{shar}}\right)^\alpha \\ 0 \end{cases} \quad (3-38)$$

3.5.2.7 Elitisme

La stratégie élitisme consiste à conserver le meilleur individu à chaque génération. Ainsi il empêche l'individu le plus performant de disparaître au cours de la sélection ou que ses bonnes combinaisons soient affectées par les opérateurs de croisement et de mutation. Après chaque évaluation de la performance des individus à une génération t donnée, le meilleur individu de la génération précédente $t-1$ est réintroduit dans la population si aucun des individus de la génération t n'est meilleur que lui. Par cette approche, la performance du meilleur individu de la population courante est monotone de génération en génération. Il apparaît que l'élitisme améliore considérablement les performances de l'algorithme génétique pour certaines classes de problème, mais engendre des dégradations pour d'autres classes, en augmentant le taux de convergences prématurées.

3.5.2.8 Sélection

La sélection a pour objectif d'identifier les individus qui doivent se reproduire. Basée sur le principe Darwinien de l'évolution, cet opérateur ne crée pas de nouveaux individus mais identifie les individus sur la base de leur fonction d'adaptation. Les individus les mieux adaptés sont sélectionnés alors que les moins bien adaptés sont écartés. Cet opérateur doit favoriser les meilleurs individus selon un critère à optimiser (minimiser ou maximiser), ceci permet de donner aux individus dont la valeur est plus grande une probabilité plus élevée de contribuer à la génération suivante [31,32].

Plusieurs opérateurs de sélection existent, parmi lesquels la sélection à la roulette, par rang et par tournoi ainsi que d'autres techniques faisant intervenir des notions de voisinage entre individus.

Sélection à la roulette

C'est la méthode la plus connue des sélections stochastiques (figure 3-12) où chaque individu de la population occupe une section de la roue proportionnellement à sa fonction d'adaptation $fitness(x_i)$, la probabilité de sélection d'un individu x_i s'écrit :

$$P_{s\acute{e}l}(x_i) = \frac{Fitness(x_i)}{\sum_{j=1}^{Popsiz} Fitness(x_j)} \quad (3-39)$$

Avec *Popsiz* le nombre d'individus de la population.

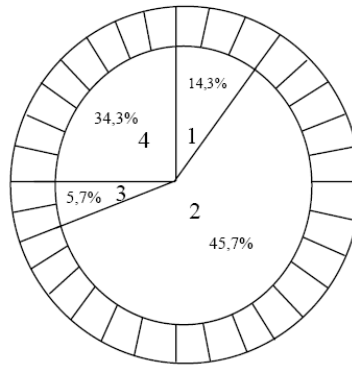


Figure 3-12 : Principe de sélection à la roulette.

A chaque fois qu'un individu doit être sélectionné, un tirage à la loterie s'effectue et propose un candidat, les individus possédant la plus grande fonction d'adaptation ayant plus de chance d'être sélectionnés.

✚ Sélection par rang

Cette sélection est divisée en deux étapes. La première ; il faut ranger les individus en ordre croissant (ou décroissant) de performance. Ensuite, une procédure de sélection similaire à la sélection par roulette est utilisée. Cette procédure permet d'attribuer une probabilité de sélection en fonction du rang, plutôt qu'avec la valeur de l'évaluation. Avec cette méthode de sélection, tous les individus ont une chance d'être sélectionnés. Cependant, elle conduit à une convergence plus lente vers la bonne solution. Ceci est dû au fait que les meilleurs individus ne diffèrent pas énormément des plus mauvais.

✚ Sélection par tournoi

Cette sélection, qui utilise des comparaisons entre individus au même titre que la sélection par rang, ne nécessite en revanche aucun tri de la population. Le principe consiste à choisir uniformément n individus et à sélectionner le meilleur. Le choix de n , qui représente la taille du tournoi, permet de faire varier la pression sélective. De cette manière on contrôle les chances de sélection des individus les plus performants par rapport aux faibles.

3.5.2.9 Croisement

Le croisement est appliqué après avoir effectué une sélection sur la population. C'est l'échange d'un certain nombre de bits entre deux chromosomes représentant deux individus de la population. En choisissant deux individus parents aléatoirement, l'opération de croisement donne naissance à deux nouveaux individus enfants [31]. Et l'opérateur a pour but d'enrichir la diversité de la population en manipulant la structure des chromosomes. Cet opérateur est contrôlée par une probabilité qui varie généralement dans l'intervalle [0.6 ; 0.8]. Il existe différents types de croisement pour un algorithme génétique, nous présentons les trois principaux types de croisement.

✚ Croisement à un point

C'est un croisement unidimensionnel dont le principe consiste à échanger les informations des bits en un point dit de coupure, le point est choisi aléatoirement sur la chaîne représentant un chromosome (voir figure 3-13).

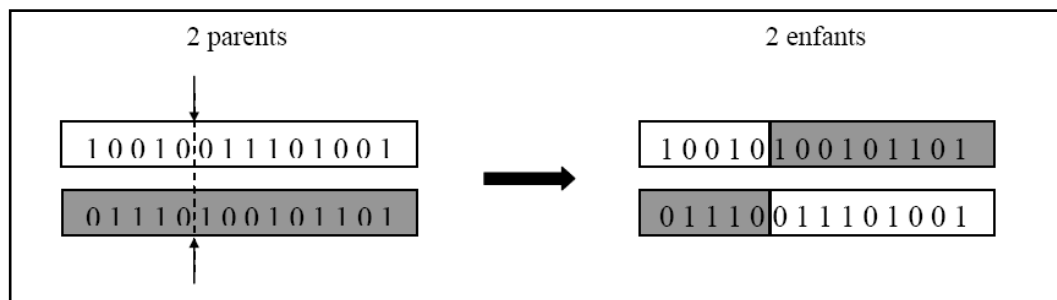


Figure 3-13 : Croisement à un point.

✚ Croisement à deux points

Le principe reste le même mais cette fois ci avec deux points de coupure, le résultat est différent du fait que les deux extrémités des chromosomes ne changent pas chez les descendants (enfants) (voir figure 3-14).

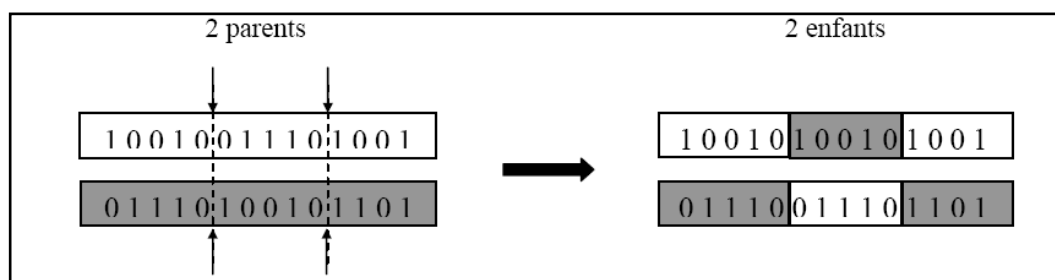


Figure 3-14 : Croisement à deux points.

🚦 Croisement uniforme

Il peut être vu, comme un croisement multipoints dont le nombre de coupure est indéterminé à priori (voir figure 3-15).

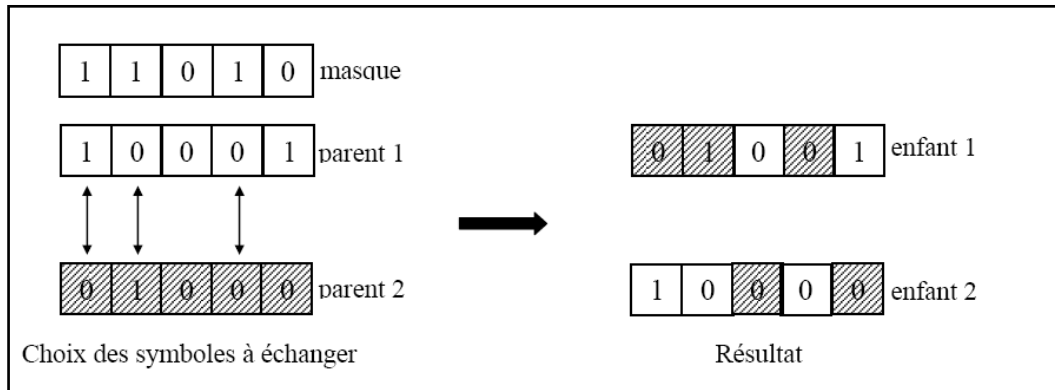


Figure 3-15: Croisement uniforme

3.5.2.10 Mutation

La mutation, (voir figure 3-16) est une modification qui intervient de manière aléatoire sur un bit (génome) d'un individu [31], elle a pour rôle de maintenir une certaine diversité dans la population et faire apparaître de nouveaux gènes. Elle intervient sur une partie suffisamment petite pour ne pas détruire les caractéristiques qui ont été sélectionnées mais suffisamment grande pour apporter des éléments nouveaux à un individu [31]. De la même manière que pour le croisement, on définit une probabilité qui contrôle l'opération de mutation, qui est de nature faible [26] de l'ordre de 0.01. Il existe différents types de mutation pour un algorithme génétique, nous présentons les trois principaux types [33].

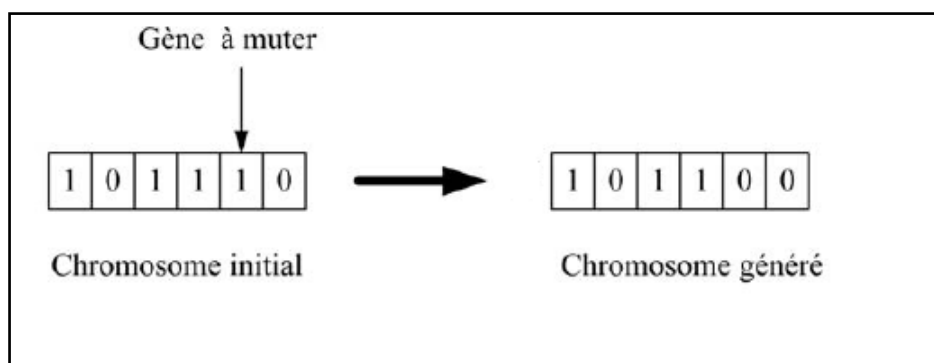


Figure 3-16 : Opérateur de mutation.

Mutation uniforme

C'est la mutation souvent utilisée dans un algorithme génétique classique. Le principe est d'altérer le contenu d'un certain nombre de gènes choisis aléatoirement.

Mutation générale

Ce type de mutation prend en compte les caractéristiques de la population dans son ensemble, généralement la diversité génétique. On effectue à chaque bit une probabilité de mutation inversement proportionnelle à sa diversité. Ce type de mutation est très efficace pour empêcher la convergence prématurée mais il ralentit aussi la convergence [34].

Mutation des composants

Il s'agit d'un opérateur local conçu spécifiquement pour les problèmes d'optimisation de forme. Le principe consiste à faire des petites modifications localisées sur les contours du domaine de conception. On utilise ce type à la fin de l'évolution afin de raffiner la forme.

Le choix de probabilités de croisement et de mutation est crucial, il doit être judicieux afin de maintenir un équilibre entre l'exploration et l'exploitation dans l'algorithme génétique. Cependant, il est difficile de maintenir cet équilibre en choisissant des probabilités fixes au cours des générations c'est pourquoi la gestion des opérateurs de croisement et de mutation peut être dynamique ou adaptative [35,36].

3.6 Avantages et désavantages des algorithmes génétiques

Le grand avantage des algorithmes génétiques est qu'ils parviennent à trouver de bonnes solutions sur des problèmes très complexes. Ils s'avèrent être des méthodes appropriées pour résoudre les problèmes d'optimisation des systèmes non linéaires et non continus [37]. Ils ne nécessitent que la connaissance des valeurs prises par la fonction à optimiser et ensuite sélectionner la meilleure. On les emploie dans les domaines où un grand nombre de paramètres entrent en jeu, et où l'on a besoin d'obtenir de bonnes solutions en quelques itérations seulement. Ils sont donc efficaces lorsque cette fonction possède plusieurs optima locaux, cas où les méthodes d'optimisation déterministe classique sont défailtantes.

Cependant, l'un des inconvénients des algorithmes génétiques est le coût en temps de calcul, qui est élevé et parfois prohibitif. Puisqu'ils manipulent plusieurs solutions simultanément. C'est le calcul de la fonction performance qui est le plus pénalisant.

Ensuite, L'un des problèmes les plus caractéristiques est celui de la *dérive génétique*, qui fait qu'un bon individu se met en l'espace de quelques générations à envahir toute la population. On parle dans ce cas de convergence prématurée et qui n'est pas forcément l'optimum attendu [37]. Les méthodes de sélection proportionnelle peuvent en particulier favoriser ce genre de dérive. Un autre problème surgit lorsque les différents individus se mettent à avoir des performances similaires : les bons éléments ne sont alors plus sélectionnés et l'algorithme ne progresse plus.

Enfin les algorithmes génétiques utilisent un grand nombre de paramètres qui doivent être fixés par l'utilisateur (taille de la population, probabilités d'application des opérateurs génétiques) et dont la mise au point de ces paramètres ne peut se faire qu'empiriquement. Pour effectuer cette mise au point, il est parfois nécessaire de procéder à de nombreux tests numériques, ce qui augmente d'autant le coût du temps de calcul.

3.7 Conclusion

Dans ce chapitre, nous avons décrit l'utilité de l'extraction de connaissance à partir des données (ECD), sa capacité d'analyser de grands volumes de données et d'être exploitée par plusieurs domaines et plus particulièrement ces dernières années en sciences et technologies. Nous avons présenté les deux grandes familles de méthodes à savoir les méthodes supervisées et non supervisées. Nous avons également donné une notion sur les réseaux de neurones et leurs applications, leurs apprentissages, et leurs utilités comme des approximateurs universels pour modéliser des systèmes complexes. Nous avons aussi décrit le principe de fonctionnement d'un algorithme génétique standard et ces différents opérateurs, et exposé les avantages et les inconvénients de ces algorithmes. La recherche bibliographique menée dans ce chapitre montre par ailleurs que les contributions de ces approches utilisées comme nouvelles techniques dans la science des matériaux fait l'objet de travaux de recherche et de réalisations existante.

Afin de mener à bien nos travaux, une base de données a été élaborée et sera décrite au chapitre correspondant. Cette collecte c'est faite à base d'une recherche bibliographique minutieuse sur différents articles des travaux théoriques et expérimentaux [38-42].

Références

- [1] Kemal Tusat Yücel, Cengiz Özel, *Advances in Engineering Software* 45 (2012) 54-60.
- [2] Hyunchul Ahn, Jae Joon Ahn, Kyong Joo Oh, Dong Ha Kim, *Expert Systems with Applications* 38 (2011) 5005-5012.
- [3] Pavel Kordik, Jan Koutnik, Jan Drchal, Oleg Kovarik, Miroslav Cepek, Miroslav Snorek, *Neural Networks* 23 (2010) 568-582.
- [4] Mst Kamrunnahar, Mirna Urquidi-Macdonald, *Corrosion Science* 53 (2011) 961-967.
- [5] Mst Kamrunnahar, Mirna Urquidi-Macdonald, *Corrosion Science* 52 (2010) 669-677.
- [6] S.C. Hui, G. Jha, *Information and Management* 38 (2000) 1-13.
- [7] Jia-Lang Seng, T.C. Chen, *Expert Systems with Applications* 37 (2010) 8042-8057.
- [8] Zhang J, Shi Y, Zhang P. Several multi-criteria programming methods for classification. *Comput. Operat. Res.* 36 (2009) 823-836.
- [9] Terzi S. Modelling the pavement present serviceability index of flexible highway pavements using Datamining. *J Appl Sci* 6(2006)193-7.
- [10] Keskin ME, Terzi Ö, Kücüksille EU, Datamining process for integrated evaporation model. *Jourl. Irrig. Drain. Eng.* 135 (2009) 39-43.
- [11] Kücüksille EU, Selbas R, Sencan A. Datamining techniques for thermophysical properties of refrigerants, *Energy Convers Manage* 50 (2009)399-412.
- [12] Adel mahamood Hassan, Abdellah Alrashdan, Mohammed T. Hayajneh, Ahmad Turki Mayyas, *Journal of Materials Processing Technology* 209 (2009) 894-899.
- [13] M.Sauget: Parallélisation de problèmes d'apprentissage par des réseaux neuronaux artificiels : Application en radiothérapie externe, Université de Franche-Comté, 7/12/2007.
- [14] Sumantra Mandal, P.V. Sivaprasad, S. Venugopal, K.P.N. Murthy, Baldev Raj, *Materials Science and Engineering A* 485 (2008) 571-580.
- [15] R. Edwin Raj, B.S.S. Daniel, *Computational Materials Science* 43 (2008) 767-773.
- [16] C.BISHOP, *Neural Network for Pattern Recognition*, Oxford University, Press 1995.
- [17] Lipo Wang, Xiuju Fu, *ACM Computing Classification* (1998) :H.2.8., I.2.
- [18] M. Yazdanmehr, S.H. Mousavi Anijdan, A. Bahrami, *Comp.Mat. Sc.* 44 (2009) 1218-1221.
- [19] Thèse Otmani I. l'Analyse discriminante et le Perceptron Multicouche, Université Mentouri Constantine (*dept. Mathématiques*) 02-06-2011.

- [20] A.S. Ramadhas, S. Jayaraj, C. Muraleedharan, K. Padmakumari, *Renewable Energy* 31 (2006) 2524-2533.
- [21] Picton P, *Introduction to neural networks* Edition Macmillan 1994 168p.
- [22] Man K.F et W.A. haling, Genetic algorithms for control and signal processing, In: *Proceeding of the 23rd International conference on Industrial Electronics, control and instrumentation, Vol.4.*New Orleans, LA, USA.pp.1541-1555.
- [23] M. Zakeri, A. Bahrami, S.H. Mousavi Anijdan, *Materials and Design* 28 (2007)2034-2039.
- [24] L.Baghli, *Contribution à la commande de la machine asynchrone, utilisation de la logique floue, des réseaux de neurones et des algorithmes génétiques*, Université Henri Poincaré, Nancy I, 14-01-1999 page 33-34.
- [25] D.E.Rumelhart, J.L.McClelland, and PDP research group. *Parallel Distributed Processing*, vol. 1-2. The MIT Press, 1986-87.
- [26] S.H.Mousavi Anijdan, A. Bahrami, H.R. Madaah Hosseini, A. Shafyei, *Materials and Design* 27 (2006) 605-609.
- [27] Y.Sun, W.D.Zeng, Y.F.Han, X.Ma, Y.Q.Zhao, *Comput. Mater. Sci.* 50 (2011) 1064-1069.
- [28] Alfonsas Misevicius, *Knowledge-Based Systems* 17 (2004) 65-73.
- [29] D.E.Goldberg, *Genetic Algorithms in Search, Optimization and machine learning*, Addison-Wesley. Reading (1989) Massachusetts.
- [30] M.Izadifar, M. Zolghadri Jahromi, *Journal of Food Engineering* 78 (2007) 1-8.
- [31] Rasit Köker, *Information Sciences* 222 (2013) 528-543.
- [32] W. Yang, Felix.T.S.Chan, V. Kumar, *Expert Systems with Applications* 39 (2012) 3081-3086.
- [33] C. Kane, *Algorithmes génétiques et optimisation topologique Ph.D. Thesis* Université de Paris VI, 1996.
- [34] C. Kane and Shoenauer, *Topological Optimum Design using Genetic Algorithms Control and Cybernetics*, 25 (5),pp.1059-1088 (1996).
- [35] W.M. Spears, *Adapting crossover in genetic algorithm*, Technical Report AIC-92-025, DC: Novel Research Laboratory, Navy center for applied Research in Artificial Intelligence, Washington, 1992.
- [36] M. Sronivas and L.M. Patnaik, *Adaptative probabilities of crossover and mutation in genetic algorithm*, *IEEE Transaction son system, Man and Cybernetics*, 24(4) (1994)656-667.
- [37] Hongming Yang, Jun Yi, Junhua Zhao, Zhao Yang Dong, *Neurocomputing* 102 (2013)154-162.
- [38] H.M.Ledbetter, *Journal of Materials Science* 20 (1985) 2923-2929.

- [39] H.M.Ledbetter, N.V.Frederick, M.W. Austin, *J. Appl. Phys.* 51(1980) 305-309.
- [40] Sudook Kim, Hassel Ledbetter, *Journal of Materials Science* 29(1994) 5462-5466.
- [41] H. M. Ledbetter, M. W. Austin, *Materials Science and Engineering*, 70 (1985) 143-149.
- [42] S.Lin, H. Ledbetter, *Materials Science and Engineering A167* (1993) 81-85.

CHAPITRE IV

APPLICATION DE LA DFT

APPLICATION DE LA DFT

4.1 Introduction

Dans ce chapitre, l'ensemble des calculs *ab-initio* présentés sont des applications de la théorie de la fonctionnelle de la densité (DFT) implémentée dans le code de calcul Vienna Ab-initio Simulation Package (VASP) [1,2], avec l'application de la méthode PAW (Projected Augmented Wave), pour lesquelles les fonctions d'onde électronique sont représentées sur des bases d'ondes planes. Tous les calculs sont effectués par l'approximation de gradient généralisé GGA [3], et la fonctionnelle d'échange-corrélation dans la forme paramétrée par *Perdew* et *Wang* (PW91) [4]. La zone de Brillouin est échantillonnée suivant la méthode de *Monkhorst-Pack* [5]. Différentes propriétés physiques (paramètres de maille, module de compressibilité) ont été calculées, leurs stabilités par rapport à l'énergie de coupure (*cut-off*) et au maillage en point k de la première zone de Brillouin (PZB) ont été aussi étudiées. Le choix de ces paramètres consiste à trouver un bon compromis entre précision et temps de calcul.

Nous avons tous d'abord effectué des calculs sur l'atome de Fer construit sur une supercell cubique à faces centrées de $2 \times 2 \times 1$ illustrée sur la figure 4-1. Par la suite, nous avons substitué quelques atomes de Fer par des atomes de Chrome (Cr) et de Nickel (Ni) pour enfin obtenir le composé ternaire $\text{Fe}_{10}\text{Cr}_3\text{Ni}_3$ ($\text{Fe}_{0.62}\text{Cr}_{0.185}\text{Ni}_{0.185}$) qui appartient à la série des aciers inoxydables dont nous avons fait une étude et calculé ses propriétés élastiques. Les deux structures étudiées sont représentées sur les figures 4-1 et 4-2.

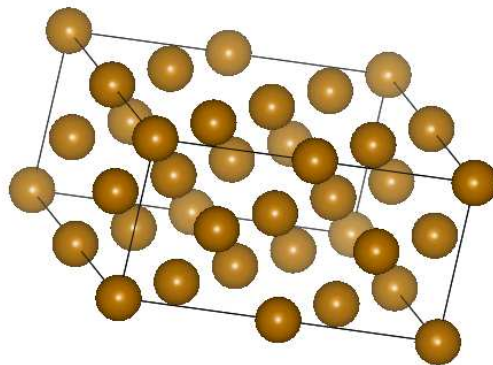
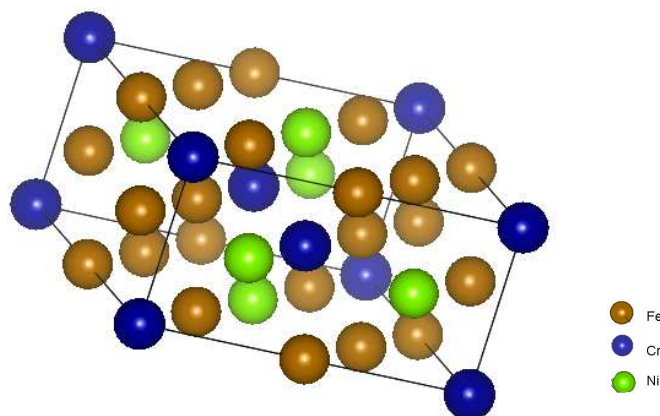


Figure 4-1 : Structure cristalline du Fer.

Figure 4-2 : Structure cristalline du $\text{Fe}_{10}\text{Cr}_3\text{Ni}_3$.

4.2 Détail de calcul

4.2.1 Equation d'état

Pour différentes valeurs du paramètre de maille prises au voisinage de la valeur expérimentale correspondante à des énergies totales différentes, les valeurs de l'énergie totale correspondantes ainsi obtenues ont été ensuite interpolées par l'équation à quatre paramètres de Birch-Murnaghan dont sa forme linéaire est donnée par l'expression suivante [6] :

$$E(V) = a + bV^{-2/3} + cV^{-4/3} + dV^{-2} \quad (4-1)$$

Avec E : énergie totale

V : volume de la maille

a , b , c et d les paramètres du graphe.

Les courbes présentées par la figure 4-3 illustrent l'évolution de l'énergie totale en fonction du volume pour une énergie de *cut-off* de 400 eV et un maillage $9 \times 9 \times 9$ de la PZB. La constante du réseau a_0 correspondant à l'état fondamental est déduite à partir du minimum de la courbe $E(V)$.

Les paramètres de maille obtenus dans ce travail sont respectivement 3.662 Å pour l'atome de Fer et 3.596 Å pour l'alliage $\text{Fe}_{0.62}\text{Cr}_{0.185}\text{Ni}_{0.185}$, objet de notre investigation. Pour l'atome de Fer, le paramètre de maille présente un parfait accord avec d'autres travaux théoriques et expérimentaux [7,8] avec une erreur relative de 0.38 %. Malheureusement pour le composé $\text{Fe}_{0.62}\text{Cr}_{0.185}\text{Ni}_{0.185}$, nous n'avons pas pu trouver dans la littérature, ni des valeurs

expérimentales ni des valeurs théoriques afin de comparer nos résultats concernant le paramètre de maille trouvé.

4.2.2 Module de compressibilité

Le module de compressibilité B qui exprime le taux de variation de la pression P en fonction du volume V est donné par l'expression suivante :

$$B = -V \frac{dP}{dV} = V \frac{\partial^2 E}{\partial V^2}_{V=V_0} \quad (4-2)$$

A pression nulle correspondant au volume V_0 de l'état d'équilibre, nous avons obtenu un module de compressibilité d'une valeur de 166.83 GPa pour l'atome de Fer avec une erreur relative de 0.87 % par rapport à la valeur expérimentale [7]. Pour l'alliage inoxydable $Fe_{0.62}Cr_{0.185}Ni_{0.185}$ (objet de notre travail) le module de compressibilité est de 152.75 GPa, sachant que la valeur expérimentale qui est de 156.7 GPa [9], l'erreur relative est de 2.52 %. Comparativement aux travaux théoriques menés par d'autres auteurs [7,8,9], nos résultats sont en bon accord.

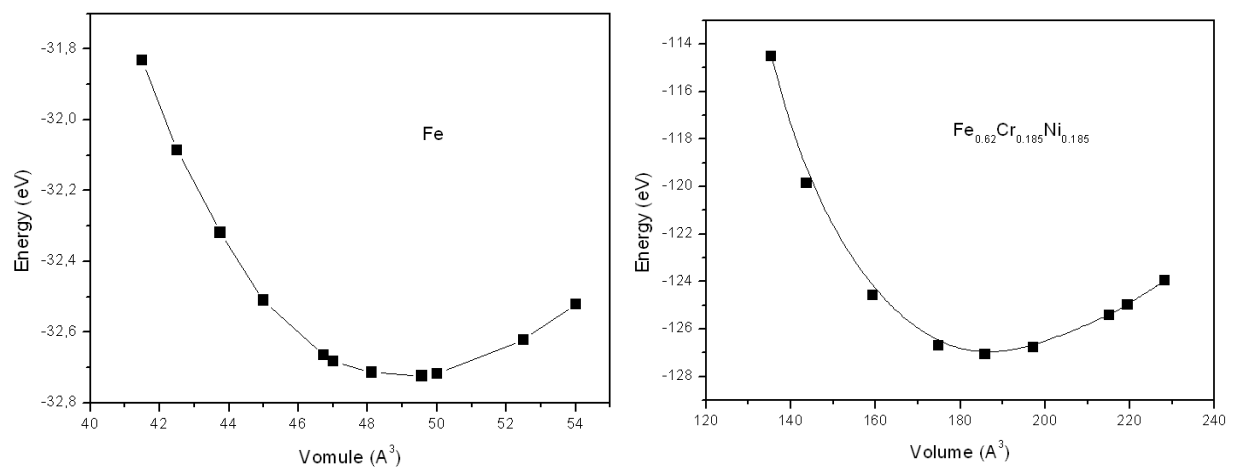


Figure 4-3 : Variation de l'énergie totale en fonction du volume de Fer et de l'alliage $Fe_{0.62}Cr_{0.185}Ni_{0.185}$.

Le tableau 4-1 donne les résultats concernant le paramètre de maille optimisé a_0 , le paramètre magnétique $M(\mu_B)$, le module de compressibilité (*Bulk modulus*) B et sa première dérivée B' . Les résultats expérimentaux et théoriques sont donnés à titre de comparaison.

Tableau 4-1 : Propriétés d'équilibre : constante du réseau a_0 , module de compressibilité B , sa première dérivée et le paramètre magnétique $M(\mu_B)$.

	a_0 (Å)	B (GPa)	B'	$M(\mu_B)$
Fe				
Nos calculs	3.662	166.83	5.69	2.67
Théorie ^a	3.632	168.00		2.54
Expérience ^b	3.648	168.3		
Fe _{0.62} Cr _{0.185} Ni _{0.185}				
Nos calculs	3.596	152.75	4.93	
Expérience ^c		156.7		

^a Ref.[7], ^b Ref.[8], ^c Ref.[9]

4.2.3 Constantes élastiques

L'élasticité d'un corps solide est une réponse sous forme d'une légère déformation quand il est soumis à des contraintes mécaniques externes. Ces contraintes sont décrites par des tenseurs qui déterminent la direction des forces et le plan sur lequel elles s'appliquent. Si la déformation est suffisamment faible, elle sera linéairement proportionnelle à la contrainte, c'est la loi de Hooke. Dans le cas tridimensionnel du solide, les contraintes σ et les déformations e s'expriment par des tenseurs d'ordre deux. Cette proportionnalité s'écrit en notation tensorielle $e_{ij} = S_{ijkl}\sigma_{kl}$, ($i, j, k, l = 1, 2, 3$), les S_{ijkl} sont des coefficients d'élasticité qui forment un tenseur d'ordre quatre. Cette relation est inversible et prend la forme $\sigma_{ij} = C_{ijkl}e_{kl}$, les C_{ijkl} étant les coefficients de rigidité. Une notation matricielle permet de simplifier les notations et la loi de Hooke qui s'écrit alors $\sigma_i = C_{ij}e_j$, ($i, j = 1, 2, \dots, 6$). Des considérations de symétrie, dont certaines dépendent du réseau de Bravais de la structure cristalline, permettent de réduire le nombre de coefficients C_{ij} indépendants. Ainsi, dans le cas particulier du réseau cubique, on reste avec trois coefficients indépendants dénommés C_{11}, C_{12} et C_{44} [10]:

$$\begin{pmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \sigma_4 \\ \sigma_5 \\ \sigma_6 \end{pmatrix} = \begin{pmatrix} C_{11} & C_{12} & C_{12} & 0 & 0 & 0 \\ C_{12} & C_{11} & C_{12} & 0 & 0 & 0 \\ C_{12} & C_{12} & C_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & C_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & C_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & C_{44} \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \end{pmatrix} \quad (4-3)$$

Cette équation (tenseur) sera utilisée par la suite dans cette partie de travail, car toutes les structures pour lesquelles on a calculé les coefficients de rigidité, ont un réseau de bravais cubique.

Ce qui nous intéresse en premier lieu dans le calcul des C_{ij} est le fait que, une fois évaluées, les C_{ij} d'un matériau permettent de juger sa stabilité mécanique. Ainsi, un solide déformé élastiquement, tout en restant dans un régime linéaire, contient une certaine énergie élastique qu'elle est possible de l'exprimer comme une densité d'énergie élastique par maille avec la forme suivante (*kittel* 1996) :

$$E = \frac{1}{2} \sum_{\lambda=1}^6 \sum_{\mu=1}^6 \tilde{C}_{\lambda\mu} e_{\lambda} e_{\mu} \quad (4-4)$$

les coefficients \tilde{C}_{ij} sont reliés aux C_{ij} par :

$$C_{ij} = \frac{1}{2} (\tilde{C}_{ij} + \tilde{C}_{ji}) \quad (4-5)$$

Cette forme quadratique de l'énergie donnée par l'expression (4-4), doit être toujours positive pour que le cristal reste stable. Si on élimine les contraintes qui causent la déformation, le cristal revient à l'équilibre en perdant l'énergie de la déformation élastique. Dans le cas contraire, si la déformation produit une énergie négative, cela provoque une diminution d'énergie totale, ce qui signifie que la structure cristalline était dans un état d'équilibre instable. De cette condition, que doit vérifier l'énergie élastique, il est possible de démontrer que les coefficients élastiques doivent satisfaire un ensemble de conditions nécessaires pour la stabilité du cristal. Ainsi, pour un réseau de bravais cubique qui a trois coefficients indépendants, les critères sont les suivants [11]:

$$\begin{aligned} C_{11} + 2C_{12} &> 0 \\ C_{44} &> 0 \\ C_{11} - C_{12} &> 0 \end{aligned} \quad (4-6)$$

En fait, il existe un grand nombre de méthodes *ab-initio* dans la littérature pour le calcul des modules d'élasticité [12-15]. Dans cette section, nous adoptons une approche qui consiste à imposer la conservation du volume de la maille sous l'effet de déformations (*volume conserving distortions methods*) [16] présentée comme suit.

✚ Calcul des constantes C_{11} et C_{12}

Pour calculer les deux constantes C_{11} et C_{12} , on applique une déformation orthorhombique à volume conservé donné par

$$I + D_0 = \begin{pmatrix} 1 + \delta & 0 & 0 \\ 0 & 1 - \delta & 0 \\ 0 & 0 & \frac{1}{1 - \delta^2} \end{pmatrix} \quad (4-7)$$

Induisant une variation de l'énergie totale donnée par l'expression :

$$\Delta E/V = 2C' \cdot \delta^2 + \theta(\delta^4) \quad (4-8)$$

Avec c' la constante élastique de Zener donnée par

$$C' = \frac{C_{11} - C_{12}}{2} \quad (4-9)$$

Par ailleurs, le module de compression pour un cristal cubique isotrope, s'écrit en fonction de C_{11} et C_{12}

$$B = \frac{C_{11} + 2C_{12}}{3} \quad (4-10)$$

✚ Calcul de la constante C_{44}

Afin de déterminer la constante C_{44} , on applique une déformation monoclinique à volume conservé :

$$I + D_M = \begin{pmatrix} 1 & \delta & 0 \\ \delta & 1 & 0 \\ 0 & 0 & \frac{1}{1 - \delta^2} \end{pmatrix} \quad (4-11)$$

La variation de l'énergie totale est donnée par l'expression :

$$\Delta E/V = 2C_{44} \cdot \delta^2 + \theta(\delta^4) \quad (4-12)$$

Pour différentes valeurs du paramètre de distorsion : $\delta = \pm 0.01n$, ($n = 0-6$), et à partir des équations (4-9), (4-10) et (4-12) les trois constantes C_{11} , C_{12} et C_{44} seront déterminés.

Les différents coefficients C_{ij} calculés sont représentés dans le tableau 4-2. On remarque que les résultats obtenus sont très proches des valeurs expérimentales [9] avec une erreur qui ne dépasse pas 2.78%. Pour tester la stabilité mécanique de cet alliage, on peut vérifier à travers les inégalités (4-6) que les conditions de stabilités sont satisfaites.

Tableau 4-2 : Constantes élastiques C_{ij} , de l'alliage $Fe_{0.62}Cr_{0.185}Ni_{0.185}$

	a_0 (Å)	C_{11} (GPa)	C_{12} (GPa)	C_{44} (GPa)
ab-initio	3.596	200.4	129.3	125.8
Expérience ^a	-	204.0	133.0	126.0
Erreur relative		1.76 %	2.78 %	0.16 %

^a Ref.[9]

Afin de vérifier la cohérence de la relaxation effectuée pour l'obtention des constantes élastiques, nous pouvons comparer le module de compressibilité (B) donné par le tableau 4-1 et celui calculé à travers les constantes C_{ij} par l'approche de Voigt et Reuss illustré dans le tableau 4-3 et qui, pour une structure cubique se réduit à l'équation (4-10).

La valeur du module de cisaillement (G) (*Shear modulus*) donnée par l'approximation de Kröner [17], peut être déterminée à partir de l'équation (4-13) en remplaçant les différentes valeurs des coefficients dans les expressions suivantes :

$$G^3 + \alpha G^2 + \beta G + \gamma = 0 \quad (4-13)$$

$$\text{avec } \left[\begin{array}{l} \alpha = (5C_{11} + 4C_{12})/8 \\ \beta = -C_{44}(7C_{11} - 4C_{12})/8 \\ \gamma = -C_{44}(C_{11} - C_{12})(C_{11} + 2C_{12})/8 \end{array} \right. \quad (4-14)$$

$$(4-15)$$

$$(4-16)$$

Les valeurs calculées à savoir le module de compressibilité (B) et le module de cisaillement (G), peuvent être utilisées afin de calculer le coefficient de Poisson $\nu = (3B - 2G)/(3B + G)$, et le module d'élasticité appelé module de Young qui est une constante reliant la contrainte de traction et la déformation d'un matériau isotrope et a pour expression $E = 9BG/(3B + G)$. Ces valeurs trouvées sont regroupées dans le tableau 4-3.

Tableau 4-3: Module de compressibilité (B), module de Young (E), module de cisaillement (G), coefficient de Poisson (ν), la pression de Cauchy (C_{12} - C_{44}), et le ratio B/G .

Alliage	B (GPa)	E (GPa)	G (GPa)	ν	C_{12} - C_{44} (GPa)	B/G
$Fe_{0.62}Cr_{0.185}Ni_{0.185}$	153.00	198.01	77.10	0.2840	3.5	1.98
$Fe_{0.62}Cr_{0.19}Ni_{0.19}^a$	156.70	198.80	77.16	0.2885	7.0	2.03

^aRef.[9]

Nous pouvons remarquer à travers le tableau 4-3 que les valeurs des différents modules et coefficients calculés présentent une bonne concordance avec les valeurs expérimentales [9]. Pour ce qui est du coefficient de Poisson, il est généralement admis que les valeurs du coefficient inférieures à 0.1 sont associées aux matériaux covalents, tandis que pour des matériaux typiquement ioniques, leurs valeurs sont proches de 0.25. Pour notre alliage ce coefficient est estimé à 0.2840 qui suggère que notre composé est un matériau à caractère ionique.

Pour un matériau défini, le module de cisaillement G est associé à une résistance d'une déformation plastique, tandis que le module de compressibilité B représente une opposition à la rupture. Par conséquent, le rapport B/G peut être considéré comme une mesure de la performance ductilité/fragilité des matériaux. Un matériau est dit ductile (capacité de changer de forme sans fracture) est caractérisé par un rapport B/G supérieur à 1.75, par contre, il est dit fragile (un changement de forme qui induit une fracture) s'il présente un B/G inférieur à 1.75 [18]. En outre, la différence entre les coefficients $C_{12} - C_{44}$ appelée pression de Cauchy, peut aussi être considérée comme un indice de la ductilité ou de la fragilité d'un matériau. Si pour un matériau, cette différence est positive le matériau est dit ductile, sinon (négative) le matériau est dit fragile. Notre investigation a montré que l'alliage $Fe_{0.62}Cr_{0.185}Ni_{0.185}$ présente un B/G égal à 1.98 qui est supérieur à 1.75 et une pression de Cauchy positive égale à 3.5 comme illustré dans le tableau 4-3, ce qui nous mène à constater que notre alliage d'acier inoxydable dévoile un caractère ductile.

4.3 Conclusion

Dans cette partie nous avons présenté les résultats des calculs menés sur l'atome de Fer et un alliage d'acier inoxydable austénitique dont la composition est $Fe_{0.62}Cr_{0.185}Ni_{0.185}$. A travers les résultats obtenus, nous avons constaté que les propriétés énergétiques et structurales sont très bien reproduites par approximation GGA, avec la fonctionnelle d'échange-corrélation dans la forme paramétrée par *Perdew* et *Wang* (PW91). Différentes propriétés physiques tels que les paramètres de maille, le module de compressibilité, et les constantes élastiques ont été déterminés. Ces derniers nous ont permis de calculer le module de cisaillement, le module de Young et le coefficient de Poisson. Il s'est avéré à travers le calcul que les résultats obtenus sont en bon accord avec ceux cités dans la littérature [7,8,9].

Dans le cadre d'étendre nos connaissances dans le domaine de la modélisation et le développement d'une nouvelle méthode qui permet la compréhension des mécanismes des propriétés mécaniques, une autre approche sera par la suite utilisée, basée sur l'application de différentes techniques du datamining pour la prédiction des propriétés élastiques.

Références

- [1] G. Kresse, J. Furthmuller, *Phys. Rev. B* 54 (1996) 11169-11186.
- [2] G. Kresse, J. Furthmuller, *Comput. Mater.Sci.* 6 (1996) 15-50.
- [3] J.P. Perdew, Y. Wang, *Phys. Rev. B* 45 (1992) 13244-13249
- [4] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, *Phys. Rev. B*, 1992, 46(11): 6671
- [5] P.O.Adebayo. K.M.Bamgbose, J.A.Olowofela, J.A.Oguntuase, G.A.Adebayo. *Physica B* 405 (2010) 4578-4581.
- [6] Shun-Li shang, Yi Wang, D.E.Kim, Zi-Kui Liu. *Computational Materials Science* 47 (2010) 1040-1048.
- [7] L. T. kong, B. X. Liu, *journal of Alloys and Compounds* 414 (2006) 36-41
- [8] C. Kittel, *Introduction to Solid State Physics*, sixth ed. Wiley, New York, 1986
- [9] A.Teklu, H.Ledbetter, S.Kim and al *Metallurgical and Mat. Sci.Trans. A* 35A (2004) 3149-3154.
- [10] M.Rajagopalan. *Physica B* 405 (2010)2516-2518.
- [11] M.Rajagopalan., S.Praveen kumar, R.Anthama. *Physica B* 405 (2010) 1817-1820.
- [12] Y. Le Page and P.Saxe, *Phys.Rev.* B63, (2001) 174103.
- [13] Y. Le Page and P.Saxe, *Phys.Rev.* B65, (2002) 104104.
- [14] M.J.Melh, *Phys. Rev.* B47, (1993) 2493.
- [15] M.J.Melh, J.E.Osburn, D.A.Papaconstantopoulous, *Phy.Rev.* B41, (1990) 10311.
- [16] C.Asker, L.Vitos, I.A.Abrikosov, *Phys Rev B* 79, 214112 (2009).
- [17] H.M.Ledbetter. *Physica* 128B (1985) 1-4.
- [18] S. F. Pugh, *Philos. Mag.* 45(1954) 823.

CHAPITRE V

APPLICATION DES TECHNIQUES

DU DATAMINING

APPLICATION DES TECHNIQUES DU DATAMINING

5.1 Introduction

Dans ce chapitre, nous avons abordé la mise en pratique des approches fondées dans le cadre du processus d'Extraction de Connaissances à partir des Données (ECD), appelé également « Datamining » qui désigne l'ensemble des méthodes ou approches destinées à l'exploration et l'analyse des données informatiques, en vue de détecter des règles, des tendances inconnues ou cachées. L'objectif du Datamining est donc de trouver des relations entre des caractéristiques (variables) suffisamment significatives et permettant ainsi d'augmenter les connaissances du domaine étudié.

Nous nous proposons par ces approches de trouver des relations de corrélations entre les propriétés physiques des composés et leurs compositions chimiques. Deux techniques seront utilisées dans notre travail. La première technique est basée sur l'application des réseaux de neurones, qui est considérée comme l'une des techniques prédictives les plus exploitées. Ils ont la capacité d'optimiser, de prédire et d'estimer des variables liées à des systèmes non-linéaires. Dans ces dix dernières années, l'utilisation des réseaux de neurone a reçu une attention croissante dans le domaine des sciences des matériaux et technologie, elles ont été appliquées par de nombreux auteurs cités dans le premier chapitre.

Néanmoins cette approche basée sur l'utilisation des réseaux de neurones tout seuls comporte certaines limitations, et de ce fait ne peut pas résoudre tous les problèmes d'optimisations. Afin de contourner ces limitations et dans le but d'améliorer les performances d'optimisations, une technique dite hybride basée sur la combinaison des réseaux de neurones avec d'autres algorithmes et notamment les algorithmes génétiques permet de soulever cet handicap. Sachant que les algorithmes génétiques sont inspirés des mécanismes de la sélection naturelle et génétique, ils utilisent les principes de la survie des individus les mieux adaptés. Ils sont considérés comme méthode stochastique d'optimisation, car ils ne nécessitent aucune hypothèse de régularité (continuité, différentiabilité), portant seulement sur la fonction à optimiser ou sur les limitations imposées, contrairement aux méthodes dites déterministes qui utilisent des informations sur la fonction à optimiser (les limitations ainsi que leur gradient). Cette dernière méthode a connu un regain d'intérêt ces

dernières années grâce au développement d'outils de calcul numérique, notamment en calcul parallèle.

Pour ne pas rester dans un cadre descriptif, nous allons présenter dans ce chapitre une application de ces approches pour une éventuelle exploitation qui consiste à une démarche de modélisation et d'optimisation. Il s'agit d'une part d'évaluer la capacité de ces approches à effectuer des tâches de simulations et de prévisions des différentes propriétés élastiques des matériaux étudiés et d'autre part de comparer leurs résultats obtenus avec ceux mentionnés dans la littérature, et ceux évalués par la méthode DFT présentée au chapitre précédent.

Dans la première partie de ce chapitre, nous présentons une application des réseaux de neurones à la modélisation de la relation composition - propriétés élastiques en utilisant plusieurs structures de réseau de neurones du type perceptron multicouche (*Multilayer Perceptron* : MLP), afin de prédire les trois différentes constantes élastiques C_{ij} de l'alliage ternaire $Fe_{0.62}Cr_{0.185}Ni_{0.185}$ cristallisant dans une structure cubique à faces centrées.

La deuxième partie de ce chapitre concerne l'application d'un modèle hybride entre le réseau de neurones et l'algorithme génétique (GA-RNA). Le réseau de neurone permettra toujours de mettre en évidence la relation de corrélation entre la composition chimique et propriétés physiques et sera élaboré de façon à devenir une fonction d'activation ou fonction fitness de notre algorithme génétique. Cette structure hybride nous permettra d'évaluer la concentration de la composition chimique optimale de l'élément manganèse (Mn) dans la composition $Fe_{(bal)}Cr_{0.185}Ni_{0.185}$ (Mn) et de prédire le module de compressibilité (B) correspondant.

Dans ce qui suit, tous nos programmes à savoir l'organigramme des réseaux de neurones, d'algorithme génétique et celui de l'hybridation (GA-RNA) ont été réalisés par nos soins sous le code de calcul MATLAB version 6.5.

5.2 Application des réseaux de neurones pour la détermination des constantes élastiques de $Fe_{0.62}Cr_{0.185}Ni_{0.185}$

Dans cette partie, notre investigation repose sur l'utilisation des réseaux de neurones comme approche, afin de trouver une corrélation entre les différentes constantes élastiques C_{11} , C_{12} et C_{44} et les éléments de composition de l'alliage du type acier inoxydable (Fe-Cr-Ni). Pour ce faire, nous avons utilisé une base de données très riche collectée par nos soins sur différents articles, illustrée dans le tableau 5-1 dont nous avons présenté les valeurs maximales et les valeurs minimales avec leurs écarts type des différentes compositions de

l'alliage Fe-Cr-Ni, et dont nous espérons qu'elle soit avantageuse et fructueuse dans l'établissement d'une bonne approche, et ensuite nous permettre de définir une relation de corrélation entre les sorties et les entrées de nos différentes structures de réseau de neurones et finalement, de pouvoir prédire les différentes constantes élastiques C_{ij} de l'acier inoxydable $Fe_{0.62}Cr_{0.185}Ni_{0.185}$.

Tableau 5-1 : Base de données utilisée dans les réseaux de neurones.

	Valeur minimum	Valeur maximum	Ecart type
Fe (Wt %)	0.10	0.969	0.2706
Cr (Wt %)	0.031	0.20	0.1738
Ni (Wt %)	0.08	0.90	0.303

Dans un premier temps, il s'agit de trouver des structures optimales en déterminant le nombre de couches cachées (*hidden layer*), le nombre de neurones pour chaque couche, et les fonctions de transfert et d'activations appliquées entre les différentes couches de neurones. Différentes architectures ont été testées et entraînées. L'erreur quadratique moyenne (*Mean Squard Error*) donnée par l'équation (5-1) a été utilisée comme critère de performance des différentes structures afin d'obtenir des résultats satisfaisants. Nous tenons à préciser que dans notre étude le nombre de noeuds formant la couche d'entrée reste fixe illustrant les différents pourcentages en poids (Wt%) à savoir le Fer, le Chrome et le nickel formant le composé ternaire Fe-Cr-Ni.

$$MSE = \left(\frac{1}{N} \sum_{i=1}^N |t_i - o_i| \right)^2 \quad (5-1)$$

Avec t_i la sortie cible (désirée)

o_i la sortie du réseau (network output).

Dans le but de supprimer l'influence des grandes valeurs sur les petites valeurs de notre base de données, toutes les valeurs ont été normalisées entre ± 0.5 et ceci par l'application de l'équation :

$$x_N = \frac{(x - x_{\min})}{(x_{\max} - x_{\min})} - 0.5 \quad (5-2)$$

Avec x la valeur originale de la base de données, x_{\min} et x_{\max} qui sont respectivement le minimum et le maximum de toutes les variables, et x_N la valeur normalisée.

Afin de trouver un meilleur modèle de réseau de neurones aidant à prédire les différents constantes élastiques, trois méthodes ont été proposées avec leurs caractéristiques et les résultats qu'ils ont obtenus.

🚩 Première méthode

Cette méthode consiste à déterminer les trois constantes élastiques C_{11} , C_{12} et C_{44} par une seule structure illustrée par la figure 5-1. Cette dernière est composée d'une seule couche cachée à 19 nœuds avec une fonction tangente sigmoïde « *tansig* » comme fonction de transfert entre la couche d'entrée et la couche cachée et une fonction d'activation de forme linéaire « *purelin* » entre la couche cachée et celle de sortie. L'évolution de la convergence de l'erreur quadratique moyenne en fonction des itérations est représentée par la figure 5-2. Nous remarquons que cette erreur tend vers une valeur de 7.873×10^{-4} et reste constante à partir de la 887^{ième} itérations.

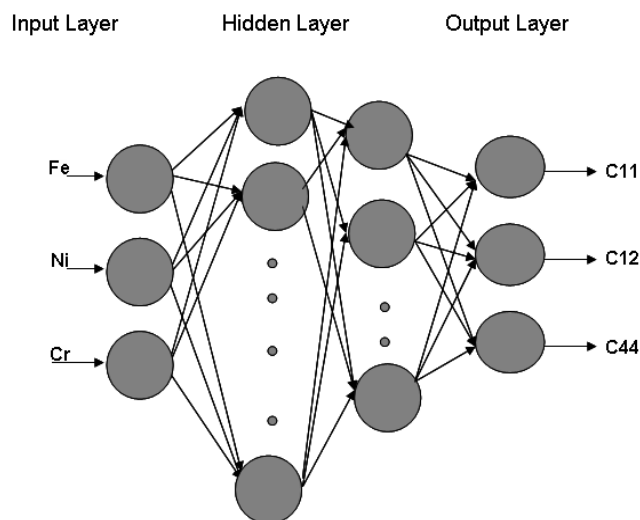


Figure 5-1: Structures de réseau de neurones à trois sorties.

Les valeurs des trois constantes élastiques prédites par cette méthode ont été mentionnées dans le tableau 5-2.

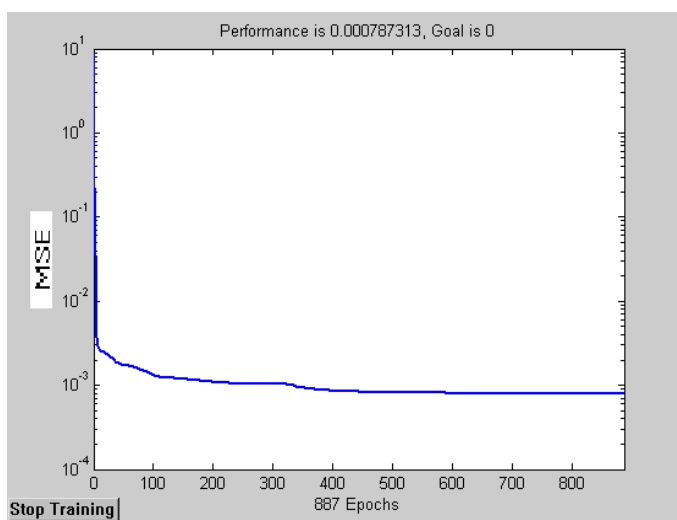


Figure 5-2 : Entraînement de convergence du réseau de neurones pour C_{11} , C_{12} et C_{44} .

Deuxième méthode

La méthode proposée consiste à la détermination des constantes élastiques séparément. Pour cela, trois modèles ont été proposés comme illustré par la figure 5-3.

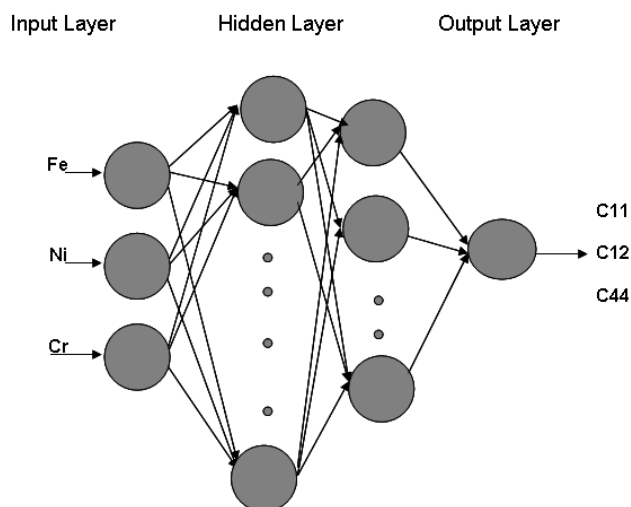


Figure 5-3 : Structures de réseau de neurones à une seule sortie (C_{11} ; C_{12} ou C_{44}).

La première constante C_{11} a été évaluée par une structure formée d'une couche cachée à 6 nœuds avec une fonction de transfert « *tansig* » et d'une fonction d'activation du type « *purelin* » pour la couche de sortie. La performance représentée par la figure 5-4 est obtenue avec une erreur quadratique moyenne de 4.001×10^{-4} après la 1000^{ème} itérations.

La deuxième constante C_{12} est estimée avec une structure à deux couches cachées, la première est composée de 6 nœuds et la seconde est composée de 10 nœuds avec une fonction

de transfert de type « *tansig* » pour chacune des deux couches et une fonction d'activation du type « *purelin* » pour la couche de sortie. La performance visualisée sur la figure 5-5 est atteinte avec une erreur quadratique moyenne de 2.174×10^{-3} et reste constante après seulement 250 itérations.

Et enfin la dernière constante C_{44} a été évaluée par un modèle similaire à celui utilisé pour la détermination de C_{11} avec une erreur quadratique moyenne de 8.515×10^{-4} atteinte après la 250^{ième} itérations (voir figure 5-6).

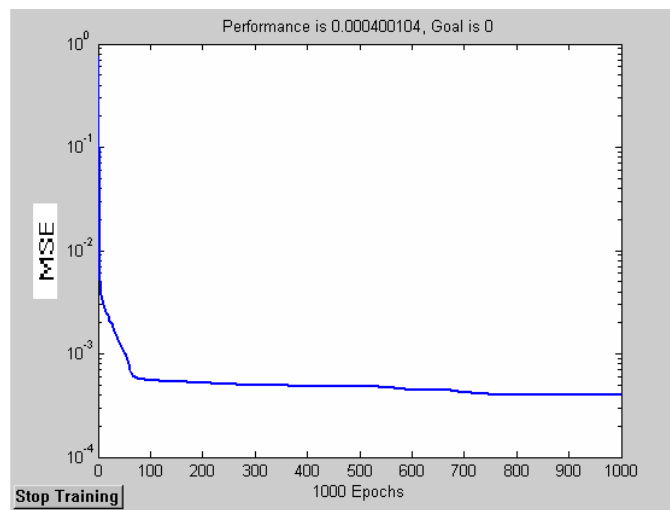


Figure 5-4 : Entraînement de convergence du réseau de neurones pour C_{11} .

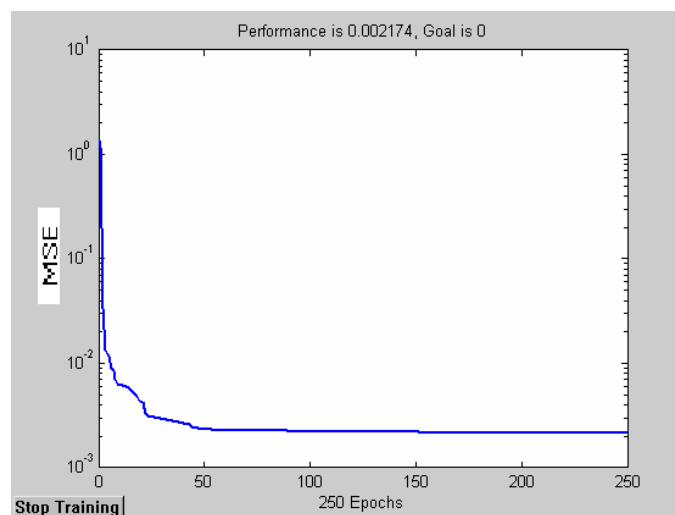


Figure 5-5 : Entraînement de convergence du réseau de neurones pour C_{12} .

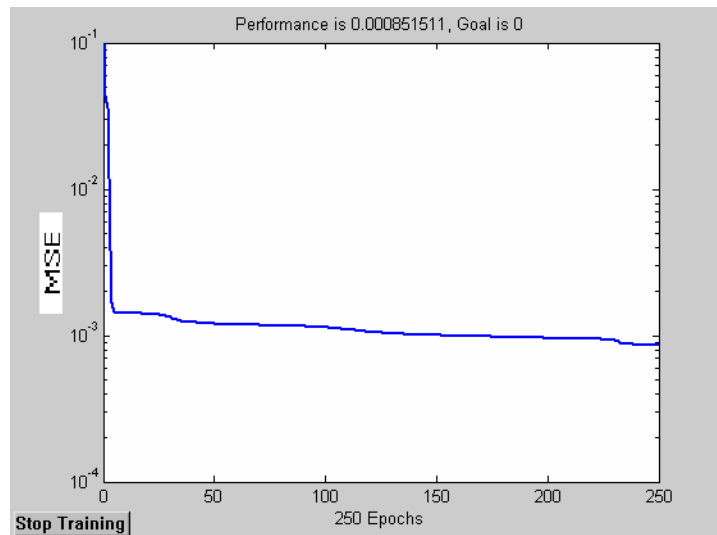


Figure 5-6 : Entraînement de convergence du réseau de neurones pour C_{44} .

Les valeurs trouvées par les différentes structures de cette méthode ont été citées dans le tableau 5-2.

Afin de confronter l'ensemble des constantes élastiques déterminées par les deux méthodes, nous avons regroupé dans le tableau 5-2 les résultats obtenus par les différentes structures définies ci-dessus. A titre comparatif, on rapporte aussi du chapitre précédent (Tableau 4-2) les résultats obtenus avec le calcul *ab-initio* et les valeurs expérimentales de la littérature.

On peut voir que l'ensemble des structures des réseaux de neurones utilisées permet de reproduire les résultats, qui à leurs tours présentent un bon accord avec les données expérimentales, et celle calculées avec la méthode *ab-initio*, en particulier les constantes élastiques déterminées par la deuxième méthode utilisant des structures à une seule sortie (C_{ij} déterminés séparément) et présentant un meilleur accord comparées aux coefficients trouvés par la première méthode avec des erreurs relatives qui sont de l'ordre de 0.49 %, 2.25 % et 0.75 % respectivement pour C_{11} , C_{12} et C_{44} .

Les figures (5-7) et (5-8) montrent une représentation des valeurs prédites par les modèles de régression donnés par les différentes structures de réseau de neurones pour la détermination de C_{11} , C_{12} et C_{44} calculés séparément en fonction des valeurs expérimentales de la base de données. Les coefficients de détermination R^2 définis par l'équation (5-3) sont 0.99, 0.95 et 0.94 pour les constantes C_{11} , C_{12} et C_{44} respectivement, indiquant ainsi un bon accord des valeurs prédites par les modèles de régression avec celles de la base de données.

$$R^2 = 1 - \left(\frac{\sum_{i=1}^N (t_i - o_i)^2}{\sum_{i=1}^N (o_i)^2} \right) \tag{5-3}$$

Avec t_i la sortie cible (désirée)

o_i la sortie du réseau (network output).

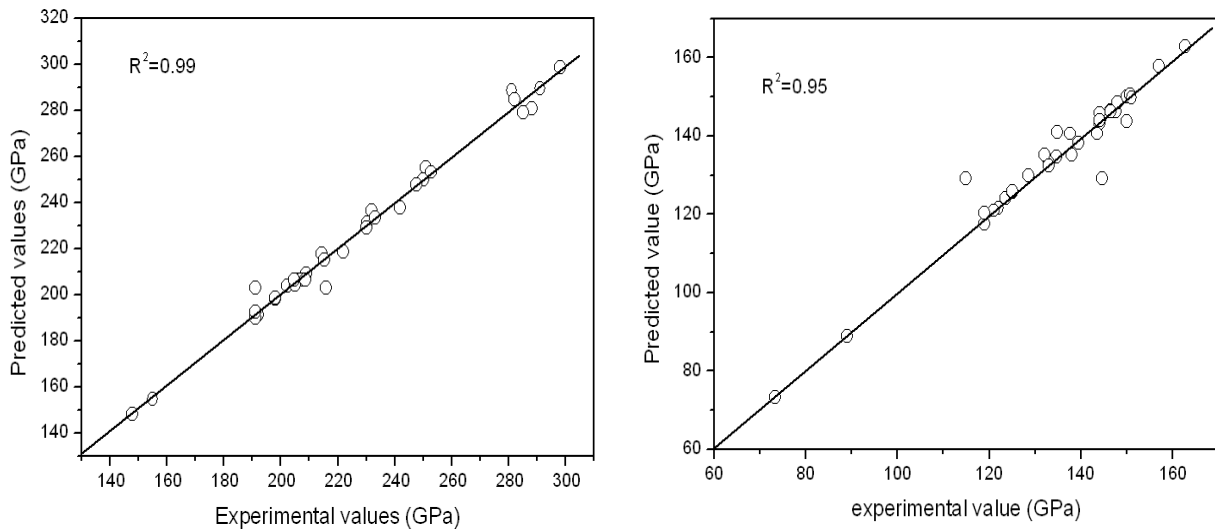


Figure 5-7 : Résultats du modèle de régression du réseau de neurones comparés avec les valeurs de la base de données pour les constantes C_{11} et C_{12} .

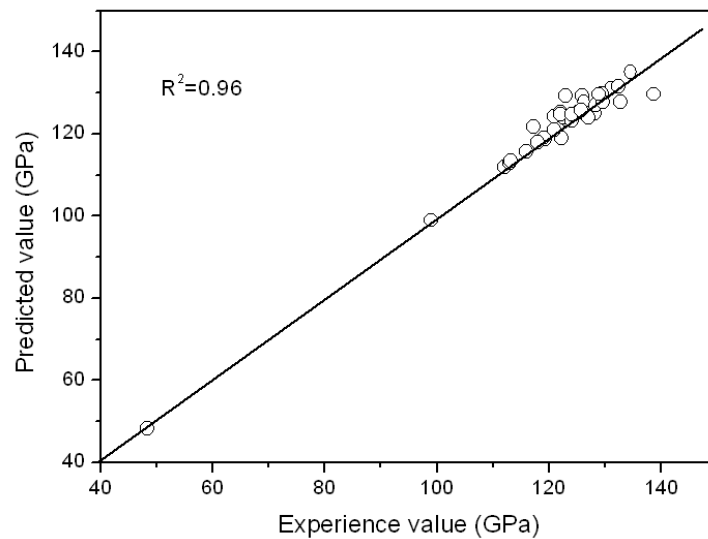


Figure 5-8 : Résultats du modèle de régression du réseau de neurones comparés avec les valeurs de la base de données pour la constante C_{44} .

Tableau 5-2: Les constantes élastiques C_{ij} avec l'erreur relative (%).

	C_{11} (GPa)	Erreur relative %	C_{12} (GPa)	Erreur relative %	C_{44} (GPa)	Erreur relative %
Expérience ^a	204.0	-	133.0	-	126.0	-
Calcul ab-initio	200.4	1.76 %	129.3	2.78 %	125.8	0.16 %
RNA (3-19-3)*	198.0	2.94	128.5	3.38	117.19	6.99
RNA (3-6-1)	203.0	0.49	-	-	126.95	0.75
RNA (3-6-10-1)	-	-	130.0	2.25	-	-
RNA (3-10-3-2)	203.87	0.064	133.86	0.64	-	-

* RNA (a-b-c) : Artificial Neural Network avec : **a** entrées, **b** nœuds sur la couche cachée et **c** sorties

^a Ref [9]

En utilisant les différentes constantes élastiques déterminées par les réseaux de neurones, le module de compressibilité (B) est calculé par l'équation (4-10), et le module de cisaillement (G) par l'approximation de Kröner (4-13), les valeurs de ces deux modules figurent dans le tableau 5-3.

Tableau 5-3 : Module de compressibilité (B) et le module de cisaillement (G) calculés.

	B (GPa)	Erreur relative %	G (GPa)	Erreur relative %
Expérience [9]	156.7	-	77.16	-
Calcul ab-initio	153.00	-	77.10	-
RNA (3-19-3)	151.66	3.21	73.16	5.18
RNA (3-6-1)	154.33	1.51	78.28	1.45
RNA (3-6-10-1)	-	-	-	-
RNA (3-10-3-2)	157.19	0.31	-	-

Notons que jusqu'ici les résultats trouvés par les deux méthodes décrites plus haut, favorisent clairement l'emploi des structures de réseaux de neurones à une seule sortie. La prédiction des constantes élastiques à travers l'application de ces structures donne des résultats très proches de l'expérience. On peut donc considérer que pour déterminer les constantes élastiques il est plus approprié de les calculer séparément.

🚩 Troisième méthode

Une autre tentative de calcul des constantes élastiques est employée en utilisant cette fois-ci une structure illustrée par la figure 5-9 composée de trois entrées qui sont les pourcentage en poids des différents éléments de composition de notre alliage (Fe, Cr, Ni), et de deux sorties C_{11} et C_{12} .

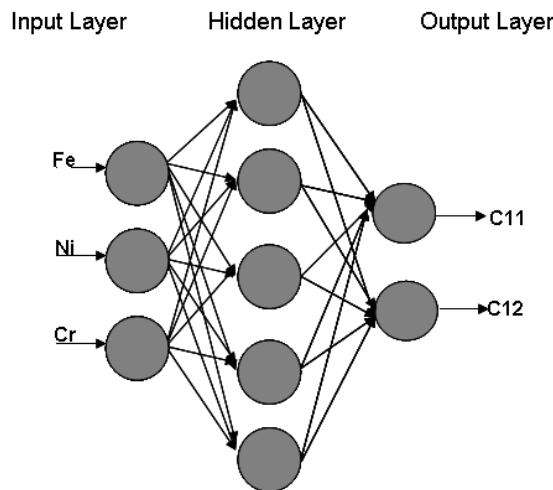


Figure 5-9: Structures de réseau de neurones à deux sorties (C_{11} ; et C_{12}).

Après plusieurs tests et entraînements, un modèle de structure a été sélectionné formé de deux couches cachées, la première composée de 10 nœuds avec une fonction de transfert de type « *tansig* », et la seconde composée de 3 nœuds avec une fonction de transfert de type logarithmique « *logsig* », et une fonction d'activation du type « *purelin* » pour la couche de sortie. La performance représentée par la figure 5-10 est atteinte avec une erreur quadratique moyenne de 1.3761×10^{-3} et reste constante après seulement 150 itérations. Les coefficients obtenus par cette structure sont listés dans le tableau 5-2.

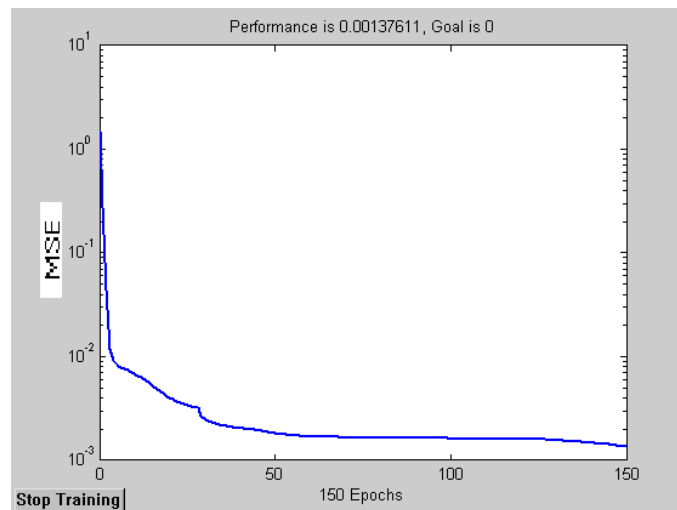


Figure 5-10 : Entraînement de convergence du réseau de neurones pour C_{11} et C_{12} .

Nous remarquons d'après le tableau 5-2 que les valeurs de C_{11} et C_{12} obtenues par la troisième méthode sont beaucoup plus précises que celles déterminées par les deux méthodes précédentes, les résultats sont en très bon accord avec les valeurs expérimentales de la

littérature [9], avec des faibles erreurs relatives de l'ordre de 0.064 % pour C_{11} et de l'ordre de 0.64 % pour C_{12} , donnant un module de compressibilité (B) d'une valeur de 157.19 GPa, qui s'approche clairement de la valeur expérimentale de 156.7 GPa avec une erreur relative de 0.31%.

Parmi les trois méthodes définies et utilisées ci-dessus afin de prédire les trois constantes élastiques de l'alliage d'acier inoxydable, la dernière tentative de calcul révèle que les résultats trouvés sont beaucoup plus précis et sont très proches des valeurs expérimentales. Cependant, pour prédire les différentes constantes élastiques C_{ij} d'une structure cubique d'un alliage il est plus judicieux de déterminer les deux constantes C_{11} et C_{12} avec un réseau de neurones à deux sorties, et la troisième C_{44} doit être calculée séparément avec un modèle à une seule sortie.

5.3 Application d'un système hybride: réseau de neurones et algorithme génétique (AG-RNA)

Dans la partie précédente, nous avons vu l'application de différentes structures de réseaux de neurones dans la prédiction des trois constantes élastiques C_{ij} , aidant à calculer le module de compressibilité (*bulk modulus*) B et le module de cisaillement (*shear modulus*) G dont les résultats sont en très bon accord avec ceux trouvés par la méthode de la DFT et ceux cités dans bibliographie. Dans cette partie et dans le même cadre de l'utilisation des techniques du Datamining, nous allons appliquer une nouvelle approche basée sur une combinaison entre le réseau de neurones et l'algorithme génétique (GA-RNA) formant ainsi un système hybride.

Dans cette nouvelle approche proposée, le réseau de neurones est utilisé comme modèle de simulation afin de trouver une relation de corrélation entre les propriétés physiques à savoir le module de compressibilité (B) et les éléments de composition de l'alliage du type acier inoxydable. Une fois l'objectif atteint, notre algorithme (RNA) sera appliqué comme fonction d'évaluation communément appelée fonction d'adaptation ou «*fitness*», qui est parmi les paramètres les plus importants de l'algorithme génétique. En faite cette fonction sert à évaluer la pertinence d'une solution par rapport à une autre. Les autres paramètres de l'algorithme génétique sont ; la sélection, le croisement et la mutation. La convergence de l'AG dépend en effet d'un choix judicieux de la méthode de sélection, ainsi que les

paramètres qui contrôlent les opérateurs de croisement et de mutation. La méthode de sélection utilisée est celle de la roulette (*wheel selection*).

Le croisement dont le principe est de produire deux nouveaux individus (enfants) en échangeant des informations entre deux nouveaux individus appelés parents. Le croisement s'effectue en deux étapes. Dans la première étape, deux individus sont choisis aléatoirement dans la population. Dans la deuxième étape, en localisant d'une manière aléatoire des points sur les deux chromosomes parents, l'opération de croisement échange le contenu (informations) de ces points entre les deux chromosomes. L'information peut être une concentration d'un des éléments chimique. Cette opération produit deux individus modifiés génétiquement qui font partie de la génération future ; ce sont les *enfants*. L'opérateur de croisement est contrôlé par un taux appelé probabilité de croisement P_c .

La mutation consiste à choisir de façon aléatoire un gène (point) au sein de l'individu et à le modifier. On définit aussi une probabilité de mutation P_m dont il est nécessaire de choisir une valeur relativement faible de manière à ne pas tomber dans une recherche complètement aléatoire qui diluerait les effets bénéfiques des principes de sélection et d'évolution.

5.3.1 Application du réseau de neurones

Dans un premier temps, notre objectif est de trouver une structure optimale, en déterminant le nombre de couches cachées, le nombre de neurones pour chaque couche, et les fonctions de transfert et d'activations appliquées par les différents neurones. Afin de modéliser une corrélation entre les différentes compositions chimiques et les propriétés physiques, une base de données sera utilisée. Cette dernière est composée de différents aciers inoxydables quaternaires (Fe-Cr-Ni-Mn) avec leurs compositions et les propriétés élastiques correspondantes qui a été collectée sur plusieurs articles, illustrée dans le tableau 5-4 dont nous présentons les valeurs maximales et minimales avec leurs écarts type des éléments d'alliage à savoir le chrome, le nickel et le manganèse (Cr, Ni, Mn).

Tableau 5-4: Base de données utilisée dans le système hybride (AG-RNA).

	Valeur minimum	Valeur maximale	Ecart type
Cr (Wt %)	14.62	23.48	1.29
Ni (Wt %)	0.118	28.08	4.74
Mn (Wt %)	0	19.97	5.33

Différentes architectures ont été testées et entraînées. L'erreur quadratique moyenne (Mean Squard Error) illustrée par l'équation (5-1) a été utilisée comme critère de performance des différentes structures afin d'obtenir des résultats satisfaisants. Toutes les valeurs de notre base de données ont été normalisées entre ± 0.5 en appliquant l'équation (5-2).

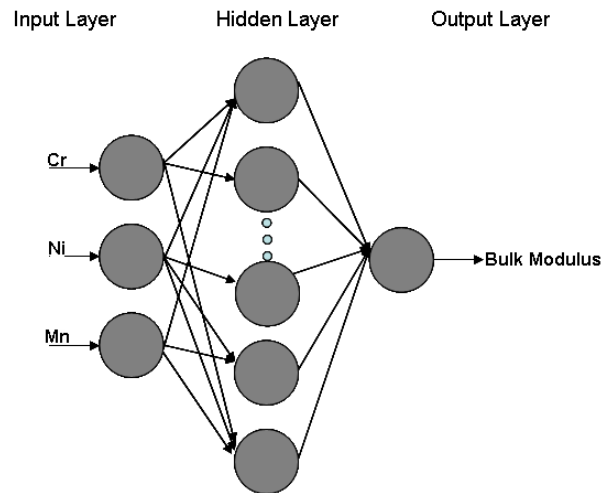


Figure 5-11 : Structures du réseau de neurones utilisées (3-couche cachée-1).

Après plusieurs essais, un modèle à trois couches a été choisi (figure 5-11), formé de trois neurones dans la couche d'entrée représentant les trois compositions chimiques Cr, Ni et Mn. Une couche cachée composée de 10 neurones et un neurone dans la couche de sortie donnant le module de compressibilité (B). La fonction de transfert entre la couche d'entrée et la couche cachée est du type «*tansig*», tandis que celle qui est entre la couche cachée et la sortie est du type «*purelin*».

La courbe donnée par la figure 5-12 montre que la performance est atteinte avec une erreur quadratique moyenne de 2.085×10^{-4} après seulement 443 itérations. La figure 5-13 expose une représentation graphique entre les valeurs prédites par le modèle de régression du réseau de neurones et les valeurs de la base de données, avec un coefficient de détermination R^2 égal à 0,9931, indiquant l'existence d'une bonne corrélation entre les valeurs de notre base de données et celles prédites. Suite à la performance atteinte par la structure RNA définie, représentée par un faible MSE et un coefficient de détermination R^2 proche de l'unité, elle peut donc être choisie comme modèle de corrélation entre les compositions chimiques et les propriétés élastiques et être capable de prédire avec un minimum d'erreur. De ce fait, elle sera utilisée comme fonction «*fitness*» pour une éventuelle optimisation dans notre système hybride (GA-RNA).

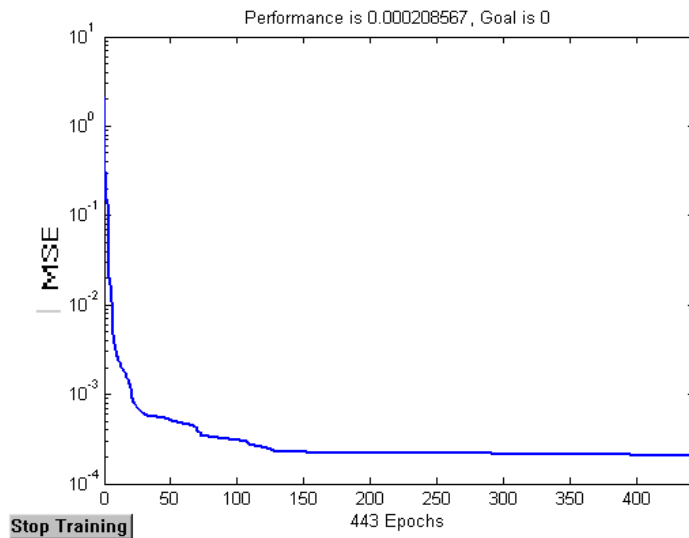


Figure 5-12 : Entraînement de convergence du réseau de neurones pour B .

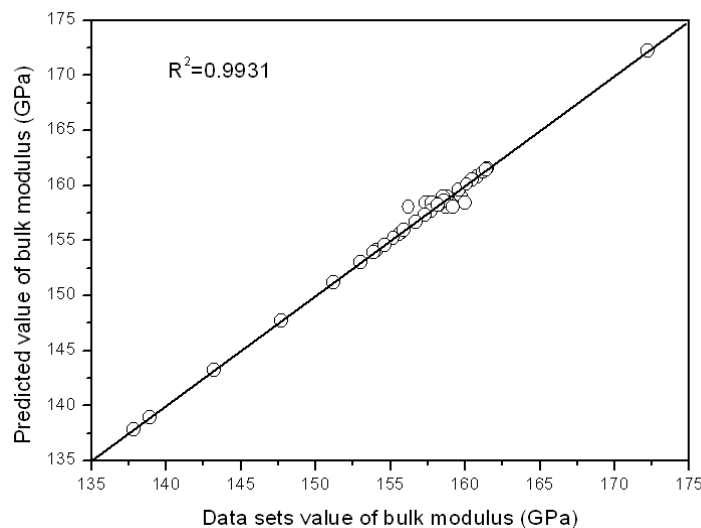


Figure 5-13: Résultats du modèle de régression du réseau de neurones pour B .

5.3.2 Application du système hybride (AG-RNA).

L'approche d'hybridation est construite par la combinaison d'un algorithme génétique comme technique d'optimisation avec le modèle du réseau de neurones étudié dans le paragraphe précédent comme fonction d'adaptation. Cette structure hybride aura pour but d'optimiser la concentration de l'élément d'alliage manganèse dans le composé quaternaire Fe-Cr-Ni-Mn et d'évaluer le module de compressibilité (B) correspondant. La figure 5-14, montre un schéma d'organigramme simple de l'hybridation de l'algorithme génétique avec le réseau de neurones (AG-RNA), et montre le fonctionnement du modèle avec la performance du RNA comme fonction fitness.

Pour mettre en œuvre l'algorithme génétique, plusieurs tests ont été effectués afin de trouver une bonne combinaison de différents opérateurs à savoir la sélection, la probabilité de croisement P_c , de mutation P_m , et la taille de la population initiale. Dans cette analyse, les concentrations des éléments chrome (Cr) et nickel (Ni) ont été maintenues fixes et ont pour valeur égale à 18.50 % en pourcentage de poids chacune, seule la concentration de l'élément manganèse (Mn) est générée par l'algorithme génétique.

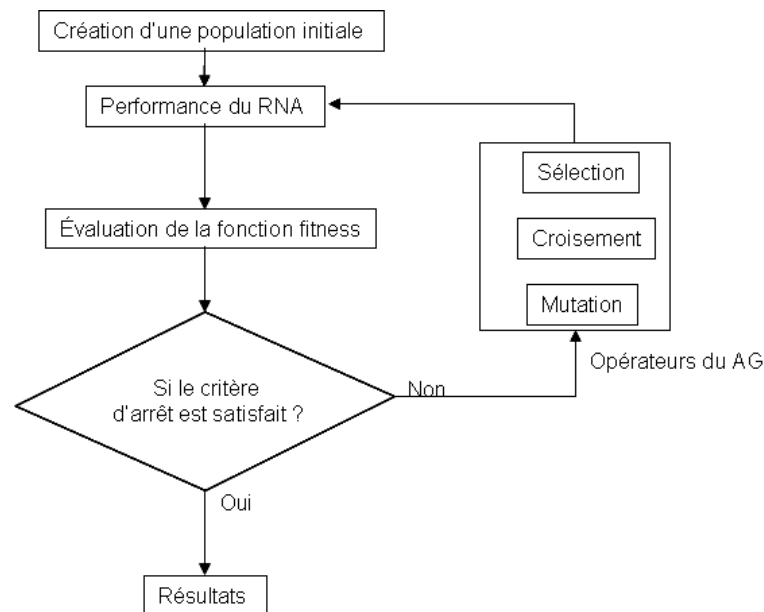


Figure 5-14 : Organigramme d'un modèle hybride (AG-RNA).

Une population initiale de 50 individus a été choisie, des probabilités $P_c = 0.8$ et $P_m = 0.001$ ont été adoptées respectivement pour l'opérateur de croisement et de mutation, une sélection à la roulette a été proposée comme opérateur de sélection. La figure 5-15 représente une évolution croissante de la fonction « *fitness* » au cours des générations, celle-ci montre que l'algorithme génétique a atteint une valeur maximale à la 100^{ième} génération. Ainsi, le résultat obtenu par notre approche d'optimisation pour la concentration de l'élément Mn dans l'alliage quaternaire Fe-Cr-Ni-Mn est de 0.1246 soit 12.46% en pourcentage de poids. Ceci indique que pour une combinaison de l'alliage composée de 18.50 % de Cr, de 18.50 % de Ni et de 12.46 % de Mn ($\text{Fe}_{(bal)}\text{Cr}_{0.185}\text{Ni}_{0.185}\text{Mn}_{0.1246}$), on obtient un module de compressibilité par la combinaison GA-RNA d'une valeur de 163.60 GPa.

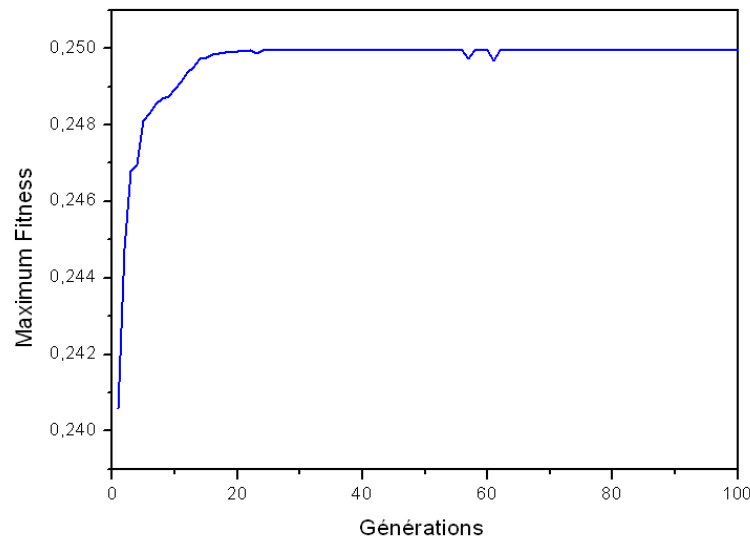


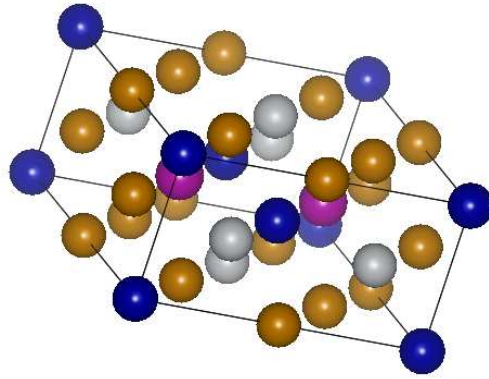
Figure 5-15 : Evolution de la fonction fitness en fonction des générations.

5.3.3 Comparaison des résultats

Afin de valider le résultat obtenu par la nouvelle approche hybride, et compte tenu de l'absence de résultats expérimentaux, il devient donc évident de le comparer à celui déterminé par le calcul *ab-initio*, basé sur la densité fonctionnelle de la théorie (DFT) implémentée dans le code de calcul Vienna Ab-initio Simulation Package (VASP) avec l'application de la méthode PAW (Projected Augmented Wave), pour lequel les fonctions d'onde électronique sont représentées sur des bases d'onde plane. Les calculs sont effectués en approximation GGA, et la fonctionnelle d'échange-corrélation dans la forme paramétrée par Perdew et Wang (PW91). La zone de Brillouin est échantillonnée suivant la méthode de Monkhorst-Pack.

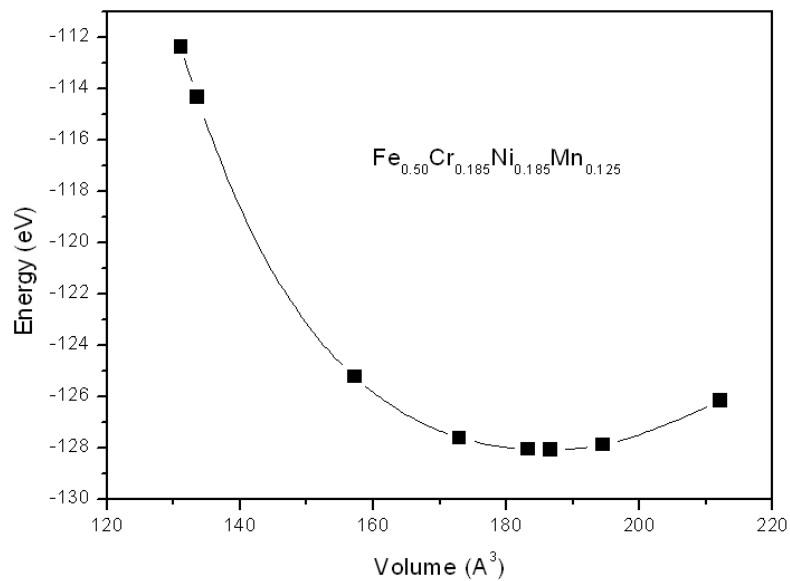
On utilise la structure de la supercell cubique $2 \times 2 \times 1$ à 16 atomes définie dans le paragraphe 4.2 du chapitre précédent, et on substitue quelques atomes de Fer, par des atomes de Nickel, de Chrome et de Manganèse pour obtenir un alliage avec des concentrations aussi proches que celles données par l'approche hybride (AG-RNA), soit 18.50 % de Cr, 18.50 % de Ni et 12.46 % de Mn. La structure obtenue est représentée par la figure 5-16, formée de trois atomes de Cr, trois atomes de Ni, et de 2 atomes de Mn (Fe ; *Bal*), $\text{Fe}_8\text{Cr}_3\text{Ni}_3\text{Mn}_2$ ($\text{Fe}_{0,5}\text{Cr}_{0,185}\text{Ni}_{0,185}\text{Mn}_{0,125}$).

Suite à la construction de la structure cristalline, différentes propriétés physiques (paramètre de maille, module de compressibilité) ont été calculées. Leurs stabilités par rapport à l'énergie de coupure (*cut-off*) et au maillage en point *k* de la première zone de Brillouin (PZB) ont été étudiées.

Figure 5-16: Structure cristalline du $\text{Fe}_8\text{Cr}_3\text{Ni}_3\text{Mn}_2$.

5.3.3.1 Equation d'état

La figure 5-17 montre l'évolution de l'énergie totale en fonction du volume pour une énergie de *cut-off* de 420 eV et un maillage $9 \times 9 \times 9$ de la PZB. La constante du réseau a_0 correspondant à l'état fondamental est déduite à partir du minimum de la courbe $E(V)$ et prend la valeur de 3.5915 Å.

Figure 5-17: Variation de l'énergie totale en fonction du volume de $\text{Fe}_{0.50}\text{Cr}_{0.185}\text{Ni}_{0.185}\text{Mn}_{0.125}$.

5.3.3.2 Module de compressibilité

A pression nulle correspondante au volume V_0 à l'état d'équilibre, le module de compressibilité qui exprime le taux de variation de la pression P en fonction du volume V peut être déduit en utilisant l'équation d'état de Birch-Murnaghan dont sa forme linéaire est

donnée par l'expression (4-1). Une valeur de 159.00 GPa est obtenue avec un B' (dérivée première par rapport au volume) égal à 4.45.

Tableau 5-5 : Module de compressibilité, calculé par la DFT et AG-RNA pour l'alliage $Fe_{0.5}Cr_{0.185}Ni_{0.185}Mn_x$.

Méthode	X = Mn (Wt%)	B (GPa)
AG-RNA	12.46	163.60
DFT	12.50	159.00

Suivant le tableau 5-5, nous remarquons que les modules de compressibilité élaborés par les deux méthodes AG-RNA et la DFT, sont du même ordre de grandeur avec une erreur relative de l'ordre de 2.89%. Il convient donc de noter que l'approche basée sur l'hybridation des réseaux de neurones avec les algorithmes génétiques choisis pour l'optimisation permet de reproduire des résultats satisfaisants qui sont en bon accord avec la méthode basée sur la DFT.

5.4. Conclusion

Dans ce chapitre, nous avons présenté les applications des techniques du Datamining sur une base de données collectée par nos soins à travers une recherche bibliographique étendue. Dans la première partie, afin de déterminer les trois constantes élastiques de l'alliage d'acier inoxydable austénitique $Fe_{0.62}Cr_{0.185}Ni_{0.185}$, les réseaux de neurones ont été employés dans le but de trouver une corrélation entre les différentes compositions (Fe, Cr et Ni) et les propriétés élastiques que nous voulons déterminer. Plusieurs structures de réseaux de neurones ont été exploitées du type perceptron multicouches. De bons résultats ont été obtenus par ces différentes structures, et particulièrement par les deux architectures à savoir la structure de type 3-10-3-2 (3 entrées, deux couches cachées la première avec 10 nœuds la seconde avec 3 nœuds et deux sorties) pour la détermination conjointe des constantes C_{11} et C_{12} , et la structure 3-6-1 (3 entrées, une couche cachée avec seulement 6 nœuds et un nœud de sortie) pour l'évaluation de la constante C_{44} seule. Les résultats trouvés sont satisfaisants et sont en bon accord comparés aux valeurs expérimentales de la base de données issue de la littérature et ceux déterminés par la méthode *ab initio*.

En second lieu, nous avons présenté une contribution originale d'une méthode d'approche qui consiste à la construction d'un système hybride formé d'une combinaison d'un algorithme génétique et d'un réseau de neurones (AG-RNA). Cette approche consiste en

l'optimisation de la concentration de l'élément Mn dans l'alliage quaternaire FeCrNi (Mn) et déduire le module de compressibilité. Après plusieurs essais, un modèle de type perceptron multicouches a été élaboré, composé de trois couches d'entrées représentant les trois compositions chimiques à savoir Cr, Ni et Mn, d'une couche cachée de 10 neurones et un neurone dans la couche de sortie donnant le module de compressibilité (B). Ensuite un algorithme génétique a été choisi, dont la fonction d'évaluation est la structure du réseau de neurones performante, avec un opérateur de sélection basé sur la roulette, une probabilité de reproduction (croisement) de 0.8 et de mutation de 0.001. Cette structure d'algorithme hybride nous a permis d'évaluer une concentration de la composition chimique de l'élément manganèse et de déterminer le module de compressibilité correspondant à l'alliage $\text{Fe}_{(bal)}\text{Cr}_{0.185}\text{Ni}_{0.185}\text{Mn}_{0.1246}$. Le résultat obtenu avec cette approche proposée a été validé et comparé à celui de la DFT avec une erreur relative estimée à 2.89%.

**CONCLUSION
ET
PERSPECTIVES**

Conclusion et perspectives

Dans notre travail, on s'intéresse à la détermination des propriétés physiques de quelques alliages d'aciers inoxydables connus par leurs excellentes propriétés mécaniques et physiques et leurs hautes résistances à la corrosion. Pour ce faire, un certain nombre d'outils de simulation et de modélisation a été élaboré, ils nous ont permis d'étudier et de déterminer ces propriétés physiques.

La théorie de la fonctionnelle de la densité (DFT), était la seule théorie longtemps utilisée dans le domaine de la chimie quantique et la science de la matière condensée, permettant de déterminer et d'étudier les différentes propriétés physiques et mécaniques des composés. Par son application pour les matériaux, elle peut apporter des informations cruciales et aider à comprendre les lois de comportement d'origine microscopique afin de guider plus efficacement le choix des matériaux pour une propriété souhaitée. L'application de la DFT implémentée dans le code de calcul Vienna Ab-initio Simulation Package (VASP) a été notre premier outil de modélisation au cours de cette thèse qui nous a permis de calculer certaines propriétés mécaniques des alliages d'aciers inoxydables, à savoir ; les constantes élastiques, le module de compressibilité et le module de cisaillement.

Ce travail se positionne en second lieu comme une contribution et un développement d'un nouveau concept d'étude et de modélisation dans le domaine de la science des matériaux à travers une extraction de connaissances des données appelées aussi « *Datamining* ». Cette contribution rentre dans le cadre du projet de « *Génomes des matériaux* », elle se situe principalement dans la modélisation des problèmes physiques et l'application des différents choix possibles des techniques et leur adaptation pour pouvoir traiter des masses de données. Afin de mener à bien notre contribution, une base de données a été élaborée contenant un ensemble d'aciers inoxydables avec leurs différentes concentrations d'éléments d'alliages (Fe, Cr, Ni, et Mn) et leurs propriétés physiques correspondantes.

Notre choix sur l'application des techniques d'exploitation s'est porté sur deux techniques supervisées : les réseaux de neurones comme méthodes de simulation des problèmes complexes connus entre les compositions chimiques des alliages et leurs différentes propriétés physiques, et les algorithmes génétiques pour leur aptitude d'optimisation. Différentes structures de réseaux neuroniques ont été développées pour la simulation et la prédiction des propriétés physiques des matériaux. Des perceptrons multicouches (MLP) ont

été utilisés afin d'établir des modèles de corrélation entre les éléments de la composition chimique de l'alliage ternaire Fe-Cr-Ni et les différentes constantes élastiques C_{ij} et d'en déduire les trois constantes : C_{11} , C_{12} , et C_{44} de $Fe_{0.62}Cr_{0.185}Ni_{0.185}$. De bons résultats ont été obtenus par ces structures et plus particulièrement par les deux architectures suivantes: la première de type 3-10-3-2 pour la détermination conjointe des constantes C_{11} et C_{12} , et la deuxième de type 3-6-1 pour l'évaluation de la constante C_{44} seule.

Une autre approche basée sur l'optimisation a été développée, construite par une combinaison hybride d'un algorithme génétique et d'un réseau de neurones (AG-RNA). Le réseau de neurones décrit par un modèle de type perceptron multicouche de forme 3-10-1 à été érigé afin de modéliser une relation de corrélation entre les éléments d'alliages d'acier inoxydable quaternaire du type Fe-Cr-Ni-Mn et le module de compressibilité B . Ce modèle donné par la structure du RNA a été utilisé aussi comme fonction d'évaluation pour notre algorithme génétique avec une population initiale de 50 individus. Une sélection à la roulette a été proposée comme opérateur de sélection, des probabilités de croisement et de mutation égales à $P_c = 0.8$ et $P_m = 0.001$ respectivement ont été adoptées. Cette hybridation (AG-RNA) ainsi définie a permis d'évaluer une concentration optimale de la composition chimique de l'élément manganèse (Mn) dans l'alliage quaternaire $Fe_{(bal)}Cr_{0.185}Ni_{0.185}Mn$ et d'en déduire le module de compressibilité B correspondant.

Une concordance entre les résultats obtenus a été remarquée lors de la confrontation entre les différentes méthodes d'approximation et de modélisation ; à savoir l'application des techniques du Datamining pour les structures de réseaux de neurones, le système hybride (AG-RNA) et la DFT. Ces résultats s'avèrent satisfaisants et en bon accord avec ceux mentionnés dans la littérature. Nous pouvons donc conclure que l'adaptation des techniques du Datamining dans le domaine de la science des matériaux pour la détermination des propriétés physiques est un outil puissant pour la modélisation et l'optimisation et ouvre des voies à des perspectives attrayantes dans le monde de la recherche scientifique.

Les réseaux de neurones peuvent être acceptés comme une technologie offrant une autre façon de simuler des problèmes complexes et d'être employés comme des modèles de prédictions. Quant aux algorithmes génétiques, ils constituent une voie prometteuse pour résoudre les problèmes d'optimisation. Par ailleurs, ces deux techniques d'approche peuvent être appliquées soit simultanément soit indépendamment entre elles dans le domaine de la

recherche scientifique et peuvent aider à l'élaboration de nouveaux matériaux aux propriétés physiques et/ou chimiques spécifiques.

Les travaux accomplis durant cette thèse ouvrent de nouvelles perspectives pour le développement constructif de l'extraction de connaissances à partir des données. Parmi celles qui nous semblent intéressantes nous pouvons citer :

- Traitement de bases de données de très grande taille.
- Utilisation d'autres techniques prédictives telles que la logique floue, les réseaux bayésiens comme des techniques de simulation.
- L'hybridation des algorithmes génétiques avec plusieurs fonctions objectives.
- L'utilisation des algorithmes génétiques à l'optimisation des structures de réseaux de neurones.

Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



(This is a sample cover image for this issue. The actual cover is not yet available at this time.)

This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at SciVerse ScienceDirect

Computational Materials Science

journal homepage: www.elsevier.com/locate/commatsci

Elastic constants of austenitic stainless steel: Investigation by the first-principles calculations and the artificial neural network approach

Kamel Benyelloul^{a,b,*}, Hafid Aourag^a

^a Applied Research Unit on Renewable Energies (URAER), BP 88, ZI, Garet Taam, Ghardaïa 47000, Algeria

^b Department of Physics, University Abou Bekr Belkaid, Tlemcen 13000, Algeria

ARTICLE INFO

Article history:

Received 12 May 2012

Received in revised form 5 September 2012

Accepted 7 September 2012

Keywords:

Austenitic stainless steel
First principles calculation
Artificial neural network

ABSTRACT

In this paper, two methods were applied to determine the different elastic constants of the face centered cubic austenitic stainless steel $\text{Fe}_{0.62}\text{Cr}_{0.185}\text{Ni}_{0.185}$. Firstly, the quantum mechanical simulation was applied based on the first principles calculations within the generalized gradient approximation (GGA) by using the efficient strain–stress method. Secondly an artificial neural network (ANN) is used based on back propagation algorithm training. ANN model has been developed for the analysis and simulation of the correlation between the elastic properties and composition. In the training model three input layers each accept the weight percentage of the alloy component (Fe, Cr and Ni), while the three different elastic constants c_{11} , c_{12} and c_{44} were employed as outputs. Different models of ANN were developed to predict the elastic constants. The performance indices such as coefficient of determination, mean square error were used to control the performance of the prediction capacity of the models developed in this study. In addition to this, elastic constants obtained from ANN models were compared with those obtained from quantum mechanical simulation and with those reported in the literature. The prediction results obtained by the two methods seem to be satisfactory.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

The austenitic stainless steels are Fe based alloys containing a minimum of 12 atomic percentage of Cr and are the most important materials of our technological society [1]. The most common austenitic grades belong to the American Iron and Steel Institute (AISI) 300 series. Their high resistance to corrosion coupled with good mechanical properties make them the material of first choice for industrial and daily applications [2]. As well as the most important used in hydrogen storage and transport [3,4]. The austenitic stainless steels are composed essentially of Fe, Cr, and Ni, and have γ -Fe face centered cubic (fcc) crystallographic structure. The mechanical properties of stainless steels depend on their chemical composition and manufacturing process [5]. Therefore, it is important to understand the correlation that exists between chemical composition and the mechanical properties. It is known that the elastic constants of materials can be used to assess certain mechanical properties such a ductility/brittleness, hardness, strength and so on [6]. Moreover, the bulk modulus, the shear modulus, Young's modulus and Poisson's ratio can be derived from

elastic constants, and are essential for determining the strength of materials [7].

Several techniques were used to calculate the elastic constants. Recently Vitos et al. [1] have used the EMTO-CPA ab initio computational tool, based on the density functional theory. They have calculated the elastic moduli of a series Fe–Cr–Ni [1]. Ganeshan et al. [8,9] have applied first-principles calculations and have determined elastic constants of Mg and binary Mg compounds using the efficient stress–strain method. For a cubic lattice, the two cubic shear modulus, $c' = (c_{11} - c_{12})/2$ and c_{44} , have been obtained from volume conserving orthorhombic and monoclinic distortions [10]. The bulk modulus B is determined by fitting an equation of state with the calculated total energy [7]. The detail of calculation can be found in Section 2. Although, that the above approach gave satisfactory results, the main drawback remains its high time execution.

In the current contribution an attempt is made to implement data-mining approach based on ANN to determine elastic constants from database related to mechanical properties of different austenitic stainless steel. The ANN has the ability to overcome searching, optimization, prediction and estimating variable related problems in complicated non-linear system [11,12]. In recent studies, ANN approach has been used by many investigators, and is receiving increasing attention in the field of materials science. Kmarunnahar and Urquidi-Macdonald [13] presented an ANN for

* Corresponding author at: Applied Research Unit on Renewable Energies (URAER), BP 88, ZI, Garet Taam, Ghardaïa 47000, Algeria. Tel.: +213 667 28 67 34.
E-mail address: Benyelloul_kamel@yahoo.fr (K. Benyelloul).

prediction of corrosion of alloy 22 (UNS N06022), Hafizpour et al. [14], presented an ANN developed to predict the compressibility of Al–SiC composite powder, Mirzadeh and Najafizadeh [15], have modeled the effect of annealing temperature and time on the reversion of strain induced martensite to austenite in the cold worker AISI 304 stainless steels alloy.

The rest of the paper is organized as follows. In Section 2, details of the theory and computational are presented. In Section 3, we presented and analyzed the result obtained by two methods. Finally, a conclusion of the present work is given.

2. Theory and computational detail

2.1. Density functional theory

The model was described by a 16-atom cubic supercell built from $2 \times 2 \times 1$ fcc. First, a Fe supercell with 16 atoms was created. Then, part of Fe atoms were randomly substituted by the Cr and Ni atoms to obtain the alloy, that $\text{Fe}_{10}\text{Cr}_3\text{Ni}_3$ ($\text{Fe}_{0.62}\text{Cr}_{0.185}\text{Ni}_{0.185}$).

First principles calculations based on the density functional theory were performed [16], and projector augmented wave (PAW) method [17,18] as implemented in the Vienna Ab-initio Simulation Package (VASP) [19,20]. We adopted the generalized gradient approximation (GGA) [21] of the Perdew and Wang (PW91) functional to describe the exchange–correlation interaction [22]. The cutoff energy was chosen as 400 eV. The k -point meshes for Brillouin zone sampling were constructed using Monkhorst–Pack scheme [23] and the $9 \times 9 \times 9$ k -points meshes is used. Spin polarized was considered in all calculation.

In order to fit the first-principles calculated total energy E as a function of volume V . We employ the four-parameter Birch–Murnaghan equation of state (EOS) with its linear form given by [24].

$$E(V) = a + bV^{-2/3} + cV^{-4/3} + dV^{-2} \quad (1)$$

where a , b , c , and d are fitting parameters. Using expressions derived from fitting parameters, we can obtain the physical properties of the alloys such a equilibrium volume (V_0), energy (E_0), bulk modulus (B) and its pressure derivative (B') [25]. Inversely, based on the fitted equilibrium properties, the fitting parameters can also be determined, see details in [24]. Following Kröner [26], the isotropic shear modulus can be estimated by the cubic equation.

$$G^3 + \alpha G^2 + \beta G + \gamma = 0 \quad (2)$$

where

$$\alpha = (5c_{11} + 4c_{12})/8, \quad (3)$$

$$\beta = -c_{44}(7c_{11} - 4c_{12})/8 \quad (4)$$

$$\gamma = -c_{44}(c_{11} - c_{12})(c_{11} + 2c_{12})/8 \quad (5)$$

Young's modulus and Poisson's ratio are calculated using the expressions.

$$E = 9BG/(3B + G) \quad (6)$$

$$\nu = (3B - 2G)/2(3B + G) \quad (7)$$

2.2. Elastic constants

The elastic constants have been calculated using volume conserving distortions. The symmetry of cubic crystal lattice reduces the 21 elastic constants to three independent elastic constants namely c_{11} , c_{12} , and c_{44} [27]. The elastic stability criteria for a cubic crystal at ambient condition are $c_{11} + 2c_{12} > 0$; $c_{44} > 0$ and $c_{11} - c_{12} > 0$ [28]. Also, it is important to maintain high accuracy

in the calculations since the total energy depends strongly on volume than on the distortions [29].

Two of these can be derived from the bulk modulus

$$B = \frac{c_{11} + 2c_{12}}{3} \quad (8)$$

And the so-called Zener's elastic constant c' ,

$$c' = \frac{c_{11} - c_{12}}{2} \quad (9)$$

The c_{11} and c_{12} elastic constants are derived from Eqs. (7) and (8). The c' elastic constant was obtained using the orthorhombic deformation,

$$I + D_0 = \begin{pmatrix} 1 + \delta & 0 & 0 \\ 0 & 1 - \delta & 0 \\ 0 & 0 & \frac{1}{1 - \delta^2} \end{pmatrix} \quad (10)$$

Yielding the following change in the total energy:

$$\Delta E/V = 2c' \cdot \delta^2 + \theta(\delta^4) \quad (11)$$

The c_{44} elastic constants can be calculated from the monoclinic distortion,

$$I + D_M = \begin{pmatrix} 1 & \delta & 0 \\ \delta & 1 & 0 \\ 0 & 0 & \frac{1}{1 - \delta^2} \end{pmatrix} \quad (12)$$

Corresponding to the energy change in

$$\Delta E/V = 2c_{44} \cdot \delta^2 + \theta(\delta^4) \quad (13)$$

In the calculations of both c' and c_{44} the distortion used were $\delta = [0.00, 0.01, \dots, 0.05]$. The elastic constants were then obtained from (linear) interpolations of the total energy as function of the square of the distortion, $\Delta E_{\text{tot}} = \Delta E_{\text{tot}}(\delta^2)$ [30].

2.3. Artificial neural networks

An artificial neural network (ANN) is a flexible mathematical model, initially developed to better understand how the mammalian brain function, which imitates the working principles of human brain [31]. The fundamental principle of ANN has been reported by several researchers [32]. There are numerous applications of ANN in data analysis, pattern recognition and adaptive control [33]. ANN models are composed of neurons connected to each others by weights. Each neuron is characterized by an input, output and activation function. In ANN literature, there exist a various sort of ANN, but the most commonly used type of ANN is the Multilayer Perceptron (MLP) [34]. MLP could operate as follows, the frontal group of neurons connected directly to the input data is called input layer while the final group of neurons delivering the results is called output layer. Other layers, as they not receive any direct input or contribute to output directly, are called hidden layers. By the architecture of an ANN, we mean the number of used neurons, their activation functions, as well as their interconnections. The term 'architecture' refers to the number of layers in the network and the number of neurons in each layer. This step is usually a trial and error procedure and various situations are required to be attempted [35].

The relationship between the inputs (x_i) and the output (y) can be written as follows [36]:

$$y = F(S) = F\left(b + \sum_{i=1}^n w_i x_i\right) \quad (14)$$

x_i is the input data, w_i the weight of the neural model, b the bias or threshold, S the activity, y the output, and F is the activation function that thresholds information.

The training of an MLP network involves finding values of the connection weights, which minimize an error function between the actual network output and the corresponding target values in the training set. Mostly back propagation training algorithm was based on back-propagation to determine the adequate weight. The coefficient of determination R^2 and mean square error (MSE) were used to evaluate the performance of the developed ANN models [37].

$$R^2 = 1 - \left(\frac{\sum_{i=1}^N (t_i - o_i)^2}{\sum_{i=1}^N (o_i)^2} \right) \quad (15)$$

$$\text{MSE} = \left(\frac{1}{N} \sum_{i=1}^N |t_i - o_i| \right)^2 \quad (16)$$

where t_i is the target output, and o_i is the network output.

3. Results and discussions

3.1. Density functional theory result

The values obtained from structure parameters optimization are used as a starting point to investigate the structure properties. In order to determine the equilibrium volume, the total energy at 0 K is calculated for six different volumes and then fitting them using four parameters Birch–Murnaghan equation of state. Subsequently, the equilibrium lattice constant, bulk modulus (B) and the first pressure derivative (B') are obtained and listed in Table 1. The bulk modulus B was found to be in the order of 152.75 GPa this value agree to a large extent with that reported in Ref. [38].

Three elastic constants namely c_{11} , c_{12} , and c_{44} have been calculated and are listed in Table 1, it can be noticed that in the case of the compounds in question, all elastic constants satisfies the criteria mentioned above. Also, bulk modulus satisfy a criterion $c_{12} < B < c_{11}$ [27], that means the compound is stable with elastic deformation. It can be seen that the present elastic constant deviates from experimental value by about 1.76%, 2.78% and 0.16% for c_{11} and c_{12} , and c_{44} respectively, this may be ascribed to the GGA-PW91 functional used in this work, which is known to underestimate the equilibrium volume.

Based on the available standard relations, Young's modulus and shear modulus are determined using the elastic constants, they are listed in Table 2. Cauchy pressure, which is the difference between c_{12} – c_{44} , serves as an indication of ductility. If the difference is positive (negative), the material is expected to be ductile (brittle). The computed Cauchy pressure was found as having a positive value, suggesting that the alloy under investigation is ductile. The B/G ratio between bulk and shear modulus may serve also as an index to qualify alloy ductility/brittleness according to Pugh criterion [39].

Table 1

Calculated lattice constant a_0 , the bulk modulus (B), its first derivative (B') and elastic constants (c_i). For iron and FeCrNi alloy compared to other theoretical and experimental data.

	a_0 (Å)	c_{11} (GPa)	c_{12} (GPa)	c_{44} (GPa)	B (GPa)	B'	M (μ_B)
<i>Fe</i>							
This work	3.662				166.83	5.69	2.67
Theory ^a	3.632				168.0		2.54
Experiment ^b	3.648				168.3		
<i>FeCrNi</i>							
This work	3.596	200.4	129.3	125.8	152.75	4.93	
Experiment ^c		204.0	133.0	126.0	156.7		

^a Ref. [41].

^b Ref. [42].

^c Ref. [38].

Table 2

Abinitio calculated Young's modulus (E), the shear modulus (G), the Poisson's ratio (ν), Cauchy pressure (c_{12} – c_{44}), and B/G ratio compared to experimental data from Ref. [38].

Alloy	E (GPa)	G (GPa)	ν	c_{12} – c_{44} (GPa)	B/G
Fe _{0.62} Cr _{0.185} Ni _{0.185}	198.01	77.10	0.284	3.5	1.98
Fe _{0.62} Cr _{0.19} Ni _{0.19} ^a	198.8	77.16	0.2885	7.0	2.03

^a Ref. [38].

The high the B/G (>1.75) ratio the ductile is the alloy and the lower of B/G (<1.75) the brittle is the alloy. Our calculation have given (B/G) ratio of 1.98 and listed in Table 2, these suggesting a mostly ductile behavior. Poisson's ratio ν are determined from the following expression (Eq. (7)), it is generally accepted that the values of Poisson's ratio smaller than 0.1 are associated with covalent material, whereas ionic materials have typical values close to 0.25 [40]. In our case, $\nu = 0.284$, and listed in Table 2, suggest a considerable ionic character in our compound.

3.2. ANN result and discussion

Through the simulation study with MATLAB programs it has been averred. The number of nodes in the hidden layer was chosen after a series of trials with networks of different structures. The number of learning data is selected from different published works [26,38,43–47]. Total datasets for model ANN is 35, among them 30 data are used in training process and 5 used in validation process.

During training of artificial neural network, the higher variable values may suppress the influence of smaller valued variables. Normalization helps to minimize the effects of magnitude variation of the data. Hence, the input and output variables are normalized between -0.5 and 0.5 . The relation used for this is given by [48]:

$$x_N = \frac{(x - x_{\min})}{(x_{\max} - x_{\min})} - 0.5 \quad (17)$$

where x is the original value from database, x_{\min} and x_{\max} are respectively the minimum and the maximum of each variable in the original data, and x_N is the normalized value.

3.2.1. Development of ANN models

In the current section, an estimation of elastic constants is proposed via two different methods using the ANN model. In the first, the elastic constants are estimated together using an ANN model through a progressive learning of three inputs data base to achieve three output targets. Secondly, each elastic constant three inputs data base is estimated by using three different training ANN models each to obtain a single output target (the ANN different structures are illustrated in Fig. 1). The chemical composition (wt.%) of alloying elements was employed as common input among all

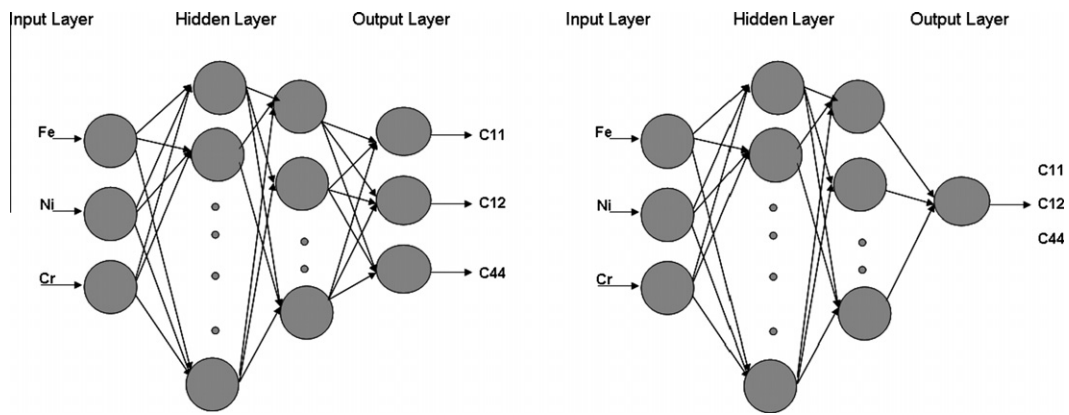


Fig. 1. Different ANN structure.

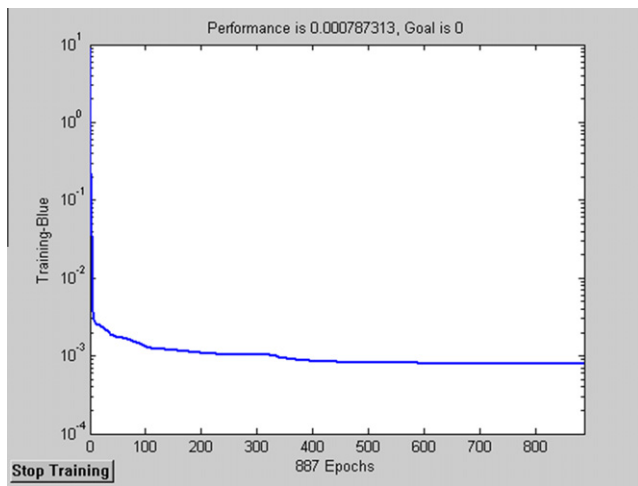


Fig. 2. Training convergence of neural network model for c_{11} , c_{12} and c_{44} .

the developed models. In order to determine the optimal architecture, different architecture artificial neural networks are trained and tested by varying the number of hidden layers, number of hidden layer nodes and activation functions. The best architecture is chosen, when the training and testing errors are minimum to produce satisfactory results [49]. After successful training and testing, the output values are checked with the actual values. The performances of the different ANN are analyzed.

The training process of the first model for three different outputs c_{11} , c_{12} , and c_{44} is illustrated in Fig. 2. After 887 epochs, the further iteration has insignificant effect on error reduction. The results of this investigation show that a MLP structure with one hidden layer, 19 nodes in the hidden layer and tangent sigmoid activation function is found to have the best performance further-

more it has a minimum mean square error (MSE) of about 7.873×10^{-4} . After that, the three elastic constants of our alloy have been predicted and are listed in Table 3.

Secondly, concerning c_{11} separately predicted. The best artificial neural network architecture which provides good training and testing is 3–6–1 (3 neurons in the input layer, 6 neurons in the hidden layer and 1 neurons in the output layer). The activation functions used are 'tansig' and the 'purelin' in the first hidden layer and output layer, respectively. The performance curve during training of the data has similar form as the former case illustrated in Fig. 2. The mean square error after 1000 epochs has achieved 4.001×10^{-4} . The model for predicted c_{12} is 3–6–10–1 (3 neurons in the input layer, 6 neurons in the first hidden layer, 10 neuron in the second hidden layer and 1 neuron in the output layer). Activation function used are 'tansig' in all layer and the 'purelin' in output layer, respectively. The neural network performance curve indicates that the convergence to mean square error of 2.174×10^{-3} is achieved in 250 epochs. The last elastic constant namely c_{44} , is predicted by the similar ANN model architecture and similar activation function for predicted c_{11} , which provide a good trained and testing result. In this case, the performance curve indicates similar convergence achieved in 250 epochs. Fig. 3 shows the regression analysis of the ANN model for single output. Coefficient of determination R^2 for c_{44} is 0.96, indicating good agreement between the experimental and the model prediction. The three ANN models are training and testing for predicted the three elastic constants separately and are listed in Table 3.

From Table 3, one can see all the ANN models developed are of acceptable accuracy within the experimental data range. It is noticeable that all elastic constants satisfies the following criteria $c_{11} + 2c_{12} > 0$; $c_{44} > 0$ and $c_{11} - c_{12} > 0$. Using c_{11} , c_{12} and c_{44} , predicted by the different ANN model, bulk modulus and shear modulus are calculated. According to this table, the elastic constants predicted by the ANN models for separately output, has a smaller relative error by about 0.49%, 2.25% and 0.75% for c_{11} , c_{12} and c_{44}

Table 3
Elastic constants c_{ij} , bulk modulus (B), shear modulus (G) and relative error (%). The first line corresponds to experimental data, the second the vasp result, and the three following to different sets of parameters.

	c_{11} (GPa)	Relative error (%)	c_{12} (GPa)	Relative error (%)	c_{44} (GPa)	Relative error (%)	B (GPa)	Relative error (%)	G (GPa)	Relative error (%)
Experiment [38]	204.0	–	133.0	–	126.0	–	156.7	–	77.16	–
VASP calculation	200.4		129.3		125.8		152.75			
ANN (3–19–3)	198.0	2.94	128.5	3.38	117.19	6.99	151.66	3.21	73.16	5.18
ANN (3–6–1)	203.0	0.49			126.95	0.75	154.33	1.51	78.28	1.45
ANN (3–6–10–1)			130.0	2.25						
ANN (3–10–3–2)	203.87	0.064	133.86	0.64	–	–	157.19	0.31	–	–

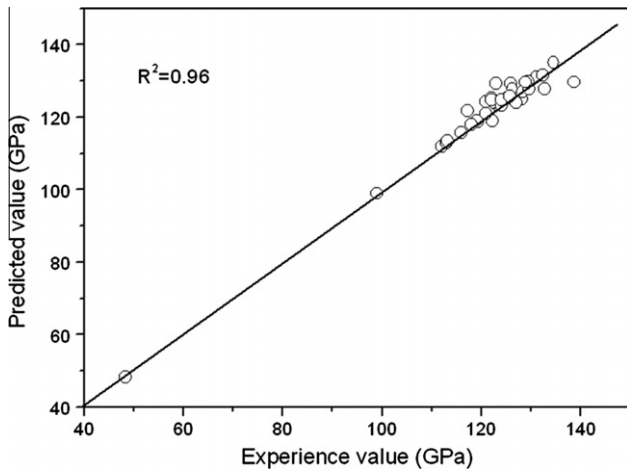


Fig. 3. ANN model simulation results compared with experimental values for c_{44} .

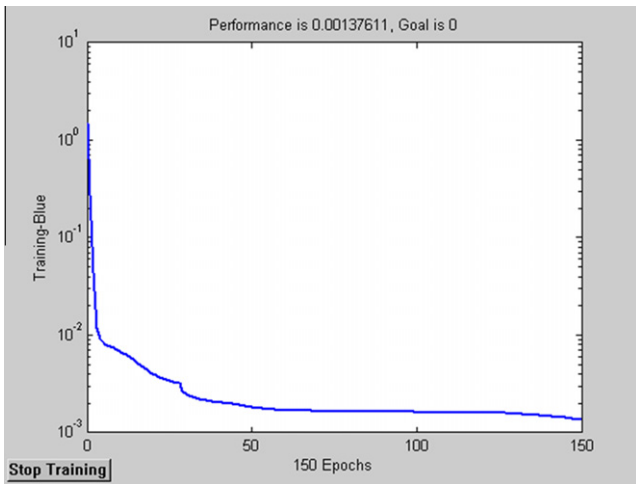


Fig. 4. Training convergence of neural network model for c_{11} and c_{12} .

respectively. In addition, B and G show good agreement with the actual data.

Another attempt of calculating elastic constants is used, ANN model with 3 input and 2 output namely c_{11} and c_{12} . After training and testing the best architecture (3–10–3–2) is selected, (3 neuron

Table 4

Predicted Young's modulus (E), the shear modulus (G), the Poisson's ratio (ν), Cauchy pressure (c_{12} – c_{44}), and B/G ratio compared to experiment.

	E (GPa)	G (GPa)	ν	c_{12} – c_{44} (GPa)	B/G
Experiment [38]	198.8	77.16	0.2885	7.0	2.03
ANNs	198.88	77.14	0.2891	6.91	2.03

in the input layer, 10 neuron in the first hidden layer, 3 neuron in the second hidden layer and 2 neuron in output layer). The activation function used is 'tansig' in the first hidden layer, and 'logsig' in the second hidden layer, 'purelin' in the output layer. The performance curve of neural network is shown in Fig. 4 which indicates that the convergence to mean square error of 1.3761×10^{-3} is archived in 150 epochs. Fig. 5 shows the regression analysis of the ANN model. The coefficients R^2 are 0.99 and 0.95 for c_{11} and c_{12} respectively, indicating good agreement between the experimental results and the predicted model. c_{11} and c_{12} are predicted and listed in Table 3.

The Table 3 shows that the two elastic constants are deviates between experimental value from about 0.064% and 0.64% for c_{11} and c_{12} respectively. This is due to the correlations that exist between the c_{11} , c_{12} and the bulk modulus for the cubic structure. Bulk modulus and shear modulus are calculated, the relative error between calculated value and experimental are 0.31% and 0.026% respectively. It must be noted, that the ANN approach models selected in our study, are consistent, and give the best results. In this scope, is proposed that the way to predict the three elastic constants of different stainless steel, from the cubic structure has been concluded by determining: c_{11} and c_{12} together then c_{44} separately. Young's modulus (E), Poisson's ratio (ν), c_{12} – c_{44} and the ratio B/G can be determined and listed in Table 4.

4. Conclusion

The purpose of this work consists mainly to compare two methods of elastic constants calculation of $\gamma\text{Fe}_{0.62}\text{Cr}_{0.185}\text{Ni}_{0.185}$ alloy. First, the elastic constants are calculated by first principles calculations within the generalized gradient approximation based on density functional formalism. The elastic properties were obtained using volume conserving strains, which is the most accurate way for calculating the elastic constants. Good agreement with experimental values has been achieved. It can be seen that the deviates results from experimental value by about 1.76%, 2.78% and 0.16% for c_{11} , c_{12} and c_{44} respectively. The calculated elastic constants have shown that $\gamma\text{Fe}_{0.62}\text{Cr}_{0.185}\text{Ni}_{0.185}$ is elastically stable on the γ

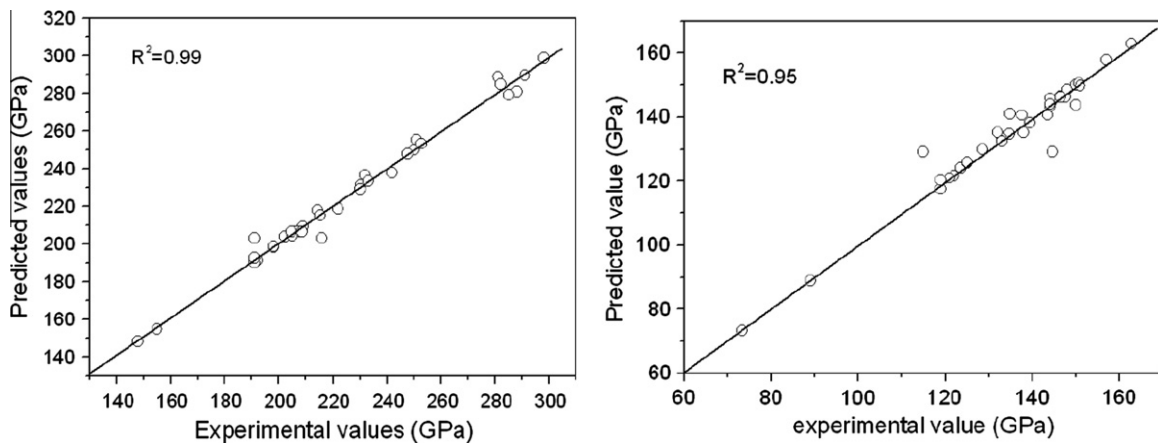


Fig. 5. ANN model simulation results compared with experimental values for c_{11} and c_{12} .

phase. Bulk modulus is determined by fitting using the 4 parameters Birch-Murnaghan equation of state, the difference is approximately 2.52% between theory and experimental value. Second, the technique of approximation using artificial neural network was successfully applied to predict the complex non-linear relationship between chemical element and elastic properties. Different models are proposed in this paper. An (3–10–3–2) ANN model structure was selected to determine the two elastic constants c_{11} , c_{12} and the c_{44} is estimate separately by the model (3–6–1) ANN. An excellent estimation result was obtained by using neural networks models. These results are in good agreement with experimental values. Through a comparative percentage errors which are found about 0.064%, 0.64% and 0.75% for c_{11} , c_{12} , and c_{44} respectively, the calculated value of the bulk modulus is 157.19 GPa, the difference is approximately 0.31%. Through the comparison of the obtained results, the ANN would be a powerful and proficient means to solve many non-linear problems in the field of material science, and overcoming the lengthy iteration time. Moreover, this approach can be used as an alternative to other experimental method thereby reducing the related cost. Although that ANN approach gave satisfactory results in case of elastic constants prediction, the DFT remains the most widely used approach in the field of material science.

References

- [1] L. Vitos, P.A. Korzhavyi, B. Johansson, Phys. Rev. Lett. 88 (2002) 15.
- [2] B. Johansson, L. Vitos, P.A. Korzhavyi, Solid State Sci. 5 (2003) 631–936.
- [3] T. Michler, Y. Lee, R.P. Gangloff, J. Naumann, Int. J. Hydrogen Energy 34 (2009) 3201–3209.
- [4] C. San Manchi, B.P. Somerday, X. Tang, G.H. Schiroky, Int. J. Hydrogen Energy 33 (2008) 889–904.
- [5] Sumantra Mandal, Mater. Sci. Eng., A 485 (2008) 571–580.
- [6] Dong Eung Kim, Shun-Li Shang, Zi-Kui Liu, Sciences 47 (2009) 254–260.
- [7] Yifang Ouyang, Xiaoma Tao, Physica B 404 (2009) 2299–2304.
- [8] S. Ganeshan, S.L. Shang, H. Zhang, Y. Wang, M. Mantina, Z.K. Liu, Intermetallics 17 (2009) 313–318.
- [9] S. Ganeshan, S.L. Shang, Y. Wang, Z.K. Liu, Acta Mater. 57 (2009) 3876–3884.
- [10] Applied parallel computing: state of the art in scientific computing: 8th international workshop, PARA 2006, LNCS 4699, 2007, pp. 510–519.
- [11] S.H. Mousavi Anijdan, A. Bahrami, H.R. Madaah Hosseini, A Shafyei Materias Des. 27 (2006) 605–609.
- [12] Kamran Dehghani, Mater. Des. 31 (2010) 2224–2229.
- [13] Mst Kmarunnahar, Mirna Urquidi-Macdonald, Corros. Sci. 53 (2011) 961–967.
- [14] H.R. Hafizpour, M. Sanjari, A. Simchi, Mater. Des. 30 (2009) 1518–1523.
- [15] H. Mirzadeh, A. Najafzadeh, Mater. Des. 30 (2009) 1518–1523.
- [16] W. Kohn, Phys. Rev. 140 (1965) 1133–1138.
- [17] G. Kresse, D. Joubert, Phys. Rev. B 59 (1999) 1758–1775.
- [18] P.E. Blochl, Phys. Rev. B 50 (24) (1994) 17953.
- [19] G. Kresse, J. Furthmuller, Phys. Rev. B 54 (1996) 11169–11186.
- [20] G. Kresse, J. Furthmuller, Comput. Mater. Sci. 6 (1996) 15–50.
- [21] J.P. Perdew, Y. Wang, Phys. Rev. B 45 (1992) 13244–13249.
- [22] J.P. Perdew, J.A. Chevary, S.H. Vosko, K.A. Jackson, M.R. Pederson, D.J. Singh, C. Fiolhais, Phys. Rev. B 46 (11) (1992) 6671.
- [23] P.O. Adebayo, Physica B 405 (2010) 4578–4581.
- [24] Shun-Li Shang, Yi Wang, D.E. Kim, Zi-Kui Liu, Comp. Mater. Sci. 47 (2010) 1040–1048.
- [25] D.E. Kim, S.L. Shang, Z.K. Liu, Intermetallics 18 (2010) 1163–1171.
- [26] H.M. Ledbetter, Physica 128B (1985) 1–4.
- [27] M. Rajagopalan, Physica B 405 (2010) 2516–2518.
- [28] M. Rajagopalan, S. Praveen kumar, R. Anthama, Physica B 405 (2010) 1817–1820.
- [29] C. Asker, L. Vitos, I.A. Abrikosov, Phys. Rev. B 79 (2009) 214112.
- [30] C. Asker, U. Kargen, L. Dubrovinsky, I.A. Abrikosov, Earth Planet. Lett. 193 (2010) 130–134.
- [31] Shouchun Wang, Xiucheng Dong, Renjin Sun, Expert Syst. Appl. 37 (2010) 4768–4771.
- [32] S.H. Mousavi Anijdan, H.R. Madaah-Hosseini, A. Bahrami, Materials Design 28 (2007) 609–615.
- [33] M. Zakeri, A. Bahrami, S.H. Mousavi Anijdan, Mater. Des. 28 (2007) 2034–2039.
- [34] J.L.R. Dias, N. Silvertre, Eng. Struct. 33 (2011) 2015–2024.
- [35] Mohsen Ostad Shabani, Ali Mozahery, Appl. Math. Model. 2011.
- [36] Z. Dahoo, Z.M. Sbartaï, A. Castel, F. Ghomari, Eng. Struct. 31 (2009) 1724–1733.
- [37] G. Dini, A. Najafzadeh, S.M. Monir-Vaghefi, A. Ebnonnasir, Comp. Mater. Sci. 45 (2009) 959–965.
- [38] A. Teklu, H. Ledbetter, S. Kim, et al., Metallurgical, Mater. Sci. Trans. A 35A (2004) 3149–3154.
- [39] S.F. Pugh, Philos. Mag. 45 (1954) 823.
- [40] A. Sekkal, A. Benzair, H. Aourag, H. Faraoun, G. Merad, Physica B 405 (2010) 2831–2835.
- [41] L.T. kong, B.X. Liu, J. Alloy Compd 414 (2006) 36–41.
- [42] C. Kittel, Introduction to Solid State Physics, sixth ed., Wiley, New York, 1986.
- [43] A. Kanrar, U.S. Ghosh, J. Phys. Chem. Solids 44 (1983) 457–462.
- [44] S.L. Shang, A. Saengdeejing, Z.G. Mei, D.E. Kim, H. Zhang, S. Ganeshan, Y. Wang, Z.K. Liu Comp. Mater. Sci. 48 (2010) 813–826.
- [45] Xiao-Qing Li, Ji-Jun Zhao, Jing-Cheng Xu. doi: 10.1007/s11467-011-0193-0.
- [46] R.E. Schramm, R.P. Reed, Metall Trans A 6A (1975) 1345–1351.
- [47] L. Vitos, Comput. Quantum Mech. Mater. Eng.: The EMTO Method Appl. (2007).
- [48] S.F. Fang, M.P. Wang, W.H. Qi, F. Zheng, Comput. Mater. Sci. 44 (2008) 647–655.
- [49] E. Edwin Raj, B.S.S. Daniel, Comput. Mater. Sci. 43 (2008) 767.



Bulk modulus prediction of austenitic stainless steel using a hybrid GA–ANN as a data mining tools



Kamel Benyelloul^{a,b,*}, Hafid Aourag^a

^a University of Tlemcen, BP.119, Tlemcen R.p 13000, Algeria

^b Division: Application of Renewable Energies in Arid and Semi Arid Environments, Applied Research Unit on Renewable Energies, EPST Development Center of Renewable Energies, URAER B.P.88, ZI, Gart Taam Ghardaia, Algeria

ARTICLE INFO

Article history:

Received 15 December 2012

Accepted 22 April 2013

Keywords:

Austenitic stainless steel
First principles calculation
Data mining
Artificial neural network
Genetic algorithm

ABSTRACT

In the current paper, the hybrid model based on genetic algorithm (GA) and artificial neural network (ANN) was used as a data mining tool to synthesize the optimal concentration of manganese (Mn). The aim is to achieve the optimal bulk modulus of FeCrNiMn austenitic stainless steel alloy. An ANN model has been developed to analyze and simulate the correlation between the elastic properties and chemical composition. The ANN training has been carried out upon three inputs, namely (Cr, Ni, and Mn), with an alloy weight percentage each, while the bulk modulus is the output target. The fitness function of GA was obtained from trained ANN model. The Mn concentration value has been obtained by the GA–ANN algorithm corresponding at the maximum bulk modulus. More ever, the given result through the GA–ANN was compared to the obtained from quantum mechanical simulation, based on the first principal calculations implemented in the Vienna Ab-initio Simulation Package (VASP). It was found that the relative error is within 2.89%. The results averred that the data mining tool based on the combination of GA–ANN is a useful, efficient, strong and adequate.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

The property of the austenitic stainless steel alloy depend on its microstructure, which in turn depends on the composition and processing of the austenitic stainless steel in a highly complex and non linear way or cannot easily be understood, a good model for predicting the proprieties of the austenitic stainless steel comes handy for controlling and/or improving the properties of the austenitic stainless steel [1]. Attempt to obtain models based on accurate experimental data in a proper manner would be a powerful tool for understanding and predicting process. In this case of situation novel approach like data mining (DM) techniques represent an assortment of tools available.

Data mining can be defined as “extraction of knowledge and information from large amount of data”. DM has recently seen an explosion of interest in many fields and subjects of applications, such a social or education science, marketing, communications, and engineering science [2–4]. DM is one of the technologies for building knowledge-based systems by inductive inference from examples [5]. Various data mining problems can be handled effectively by soft computing techniques. These techniques are fuzzy logic, neural network and genetic algorithm.

* Corresponding author at: University of Tlemcen, BP.119, Tlemcen R.p 13000, Algeria. Tel.: +213 667 286 734.

E-mail address: benyelloul_kamel@yahoo.fr (K. Benyelloul).

Artificial neural network (ANN) has popularity in solving several problems and technical problems that involve prediction, and have a wide ranging usage area. Is one of the most important data mining techniques [6]. ANN has become a very powerful and practical method to model very complex non-linear system. The ANN is widely applied in optimization, classification and prediction [7]. Recently, ANN approach has been used by many investigators. Edwin Raj and Daniel [8] are developed ANN models for the simulation of compressive proprieties of closed-cell aluminium foam. Ozerdem and Kolukisa [9] have employed ANN to predict the mechanical properties of Cu–Sn–Pb–Zn–Ni cast alloys. Hayajneh et al. [10] are used the neural network to predict wear loss quantities of some aluminium–copper–silicon carbide composite materials. But ANN may not solve all problems in the real world and it has some limitations. In order to prevent these disadvantages, researches have proposed many alternatives [11]. The Hybridizing ANN with other approaches in the search for better performance. Genetic algorithm (GA) is efficient alternative. The combined GA–ANN algorithms has a great potential to handle problems such as optimization in complicated nonlinear systems [12,13]. In recent years, much attention has been focused on the hybrid GA and ANN approach. A number of related studies have been performed in the past few decades. In the early studies the hybrid GA–ANN has widely used, and is receiving increasing attention in the materials science. Song and Zhang [13] used the GA–ANN to search the optimum heat-treatment adapted for

7175 aluminium alloys. Yazdanmehr et al. [14] GA–ANN have been used to find the optimal conditions for achieving the minimum coercivity and crystallite size of $\text{Fe}_{85}\text{Si}_{15}$. Recently Sun et al. [15] have been employed the hybrid GA–ANN to synthesize the optimum chemical composition for optimize the mechanical properties for TC11 titanium alloy.

Through the current work, ANN was used to find the relationship between different chemical composition (Cr, Ni and Mn) and elastic property. Further, the above model is used as fitness function of GA algorithm. A coupled GA–ANN optimization methodology was used to find an appropriate chemical composition of manganese that leads the highest bulk modulus in FeCrNiMn alloy; and then, the results were compared between quantum mechanical simulation based in density functional theory (DFT) and combined GA–ANN algorithm.

2. Different algorithm development structures

The development detail of different prediction algorithms has been investigated through the following.

2.1. Artificial neural network

The artificial neural network (ANN) is a computational structure inspired by a biological neural system [16]. A detailed description has been mentioned in the published paper [17]. As it is described, ANN model is composed of neurons connected to each other by some weighted links [18]. Each neuron is characterized by the input, output and activation function. Among different types of ANNs, multilayer perceptron (MLP) neural networks are quite popular and used in the present work [19]. The MLP could operate as follows: the frontal group of neurons connected directly to the input data is called input layer, while the final group of neurons delivering the results are called output layer. Other layers, as they not receive any direct input or contribute to output directly, are called hidden layers. The relationship between the inputs (x_i) and the output (y) can be written as follows [20].

$$y = F(S) = F\left(b + \sum_{i=1}^n w_i x_i\right) \quad (1)$$

where x_i is the input data, w_i the weight of the neural model, b the bias or threshold, S the activity, y the output, and F is the activation function that thresholds information.

The performance of the ANN was statistically measured by mean squared error (MSE) and the coefficient of determination obtained as follows [18,21]:

$$\text{MSE} = \left(\frac{1}{N} \sum_{i=1}^N (t_i - o_i)\right)^2 \quad (2)$$

$$R^2 = 1 - \left(\frac{\sum_{i=1}^N (t_i - o_i)^2}{\sum_{i=1}^N (o_i)^2}\right) \quad (3)$$

where t_i is the target output, and o_i is the network output.

2.2. Genetic algorithm

Genetic algorithm (GA) is one of classic population-based robust search and global optimization methods, which is developed by John Holland and his colleagues at the University of Michigan [15]. Is becoming an important tool for optimizing functions [19,22]. Based on the evolutionary process of biological organisms in nature the algorithm solves optimization problems imitating nature in the way it has been working on the evolution of life for millions of years [20,23]. These algorithms are used to solve linear

and non-linear optimization problems by exploring all regions of state space and exploiting potential areas through mutation, crossover and selective operations applied to individuals in the population [24,25]. After beginning with randomly generated population, the *selection* operator chooses chromosomes from the current population based on fitness value of the individuals [25]. The *crossover* concerns two selected individuals (parents) randomly but proportional to their fitness function values that exchange parts of their genomes to form two new individuals (children) [26]. Generally one-point or two-point crossover is commonly used as crossover method [22]. Finally, *mutation* operator process follows the crossover procedure [27], the role of mutation process is to provide the diversity in the population, and consist merely randomly selects one individual from the parent population and change its internal representation and puts it in the child's population [21]. Since mutation rate is very small in nature evolution, the probability at which the mutation operator is applied is set to a very low value [19]. Both of the crossover and mutation rate should be carefully adjusted to improve search performance [21]. Once all operators are performed on the initial population, a new population is generated. This new population is genetically better at the previous and has a greater chance of survive for the given problem condition. This process continues until an optimal solution is found or reached the maximum number of generations which is pre-set by the users.

The current step outlines some of the basics of genetic algorithms. The three most important aspects of using genetic algorithms are: (1) definition of the objective function, (2) definition and implementation of the genetic representation, and (3) definition and implementation of the genetic operators. Once these three have been defined, the generic genetic algorithm should work fairly well [28].

2.3. Integration of GA and ANN

The combined GA–ANN model which has been used in this investigation is shown in the flowchart presented in Fig. 1. An initial population is generated randomly composed by different concentration of Mn in FeCrNiMn composition. Each individual is evaluated in the initial population by the fitness function, based on the trained ANN model. Then, the next generation is created using the genetic operators (*crossover* and *mutation*). Once all main operators are performed on the initial population, the new gener-

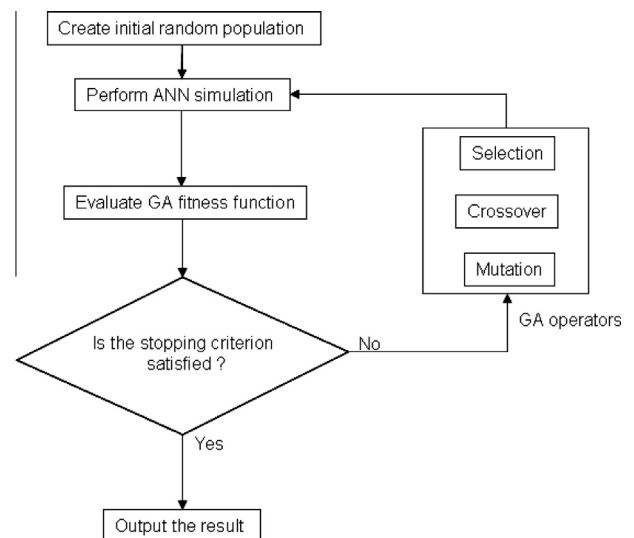


Fig. 1. Flowchart of GA–ANN model.

ation has been developed and has genetically superior to the previous one. Furthermore it was better chance to survive for the given problem condition. This algorithm is repeated until the maximum generation number or population convergence is reached [27,29].

2.4. Description of the DFT theory

In quantum mechanical simulation theory, the first principal calculation based on the DFT was used. The projector augmented wave (PAW) method [30,31] as implemented in the Vienna Ab-initio Simulation Package (VASP) [32,33]. The generalized gradient approximation (GGA) has been adopted [34] of the Perdew and Wang (PW91) functional to describe the exchange–correlation interaction [35]. The k -point meshes for Brillouin zone sampling were constructed using Monkhost-Pack scheme [36]. In the aim to fit the first-principles calculated total energy E as a function of volume V . The 4-parameter Birch–Murnaghan equation of state (EOS) has been employed with its linear form, which is detailed in the publish [37].

$$E(V) = a + bV^{-2/3} + cV^{-4/3} + dV^{-2} \quad (4)$$

where a , b , c , and d are the fitting parameters.

3. Results and discussions

The initial data used for this study have been selected from different published works [38–42]. In this investigation, total data sets are 43 samples. Among them 35 data were used in the training of the ANN structure, as described in Fig. 2 while the remaining eight data were used the generalization capability of ANN. The maximum and minimum values as well as the standard deviation of the ANN inputs are shown in Table 1.

3.1. ANN result

Based on the above algorithm, for the process simulation and optimization has been developed using MATLAB programs. In order to develop the optimum architecture of ANN, the number of hidden layer was chosen after a series of trials with networks of different structures [27]. In addition, the number of nodes in hidden layer was also determined after series of trials and tested by the calculated maximum absolute error (MAE) given by following equation [43].

$$MAE = \sum |x - y|/n \quad (5)$$

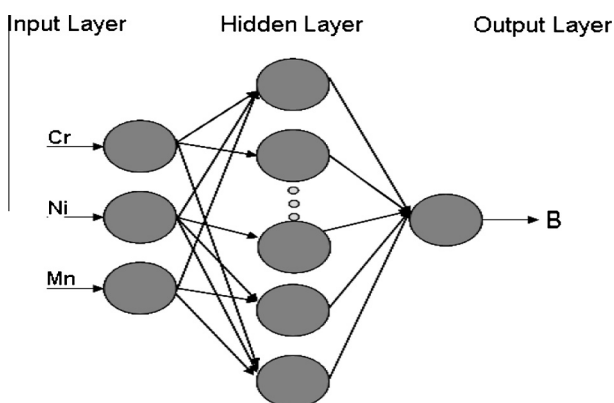


Fig. 2. Schematic representation of multilayer perceptron (MLP).

Table 1

Input data used to develop the neural network model for FeCrNiMn stainless steel alloy.

	Minimum value	Maximum value	Standard deviation
Cr (wt%)	14.62	23.48	1.29
Ni (wt%)	0.118	28.08	4.74
Mn (wt%)	0	19.97	5.33

where $x = X - X'$; X is the target output and X' is the mean X , $y = Y - Y'$; Y is the network output and Y' is the mean of Y .

Fig. 3 shows the neuron number effect in the hidden layer on the MAE, which showed the optimum number of neurons that gives the minimum error in the hidden layer, is 10.

During training of artificial neural network, the higher variable values may suppress the influence of smaller valued variables. Normalization helps to minimize the effects of magnitude variation of the data. Hence, all the input and output data were normalized to ± 0.5 according to the following equation [44].

$$x_N = \frac{(x - x_{\min})}{(x_{\max} - x_{\min})} - 0.5 \quad (6)$$

where x is the original value from database, x_{\min} and x_{\max} are respectively the minimum and the maximum of each variable in the original data, and x_N is the normalized value.

Three ANN layers were used. The inputs are the chemical composition Cr, Ni and Mn of austenitic stainless steel alloys and the output is the elastic properties. The selected ANN architecture that provides good training and testing is composed of three input nodes, 10 nodes in the hidden layer and one output node. Transfer function between the input layer and hidden layer is 'tansig', while the activation function between the hidden layer and the output is 'purelin'. The training of the neural network presented in Fig. 4, illustrated the performance curve, which averred that a mean square error (MSE) about 2.085×10^{-4} is achieved, after 443 epochs. Fig. 5 shows the regression analysis of the ANN model and coefficient of determination R^2 for B is 0.9931, which indicates a good agreement between the dataset and model prediction. The acquired performances indicate that the developed ANN could be used for prediction with a minimum error; more ever it could provide an accurate relationship between the chemical composition and the bulk modulus. The simulation function based on the ANN, was used as a fitness function in GA.

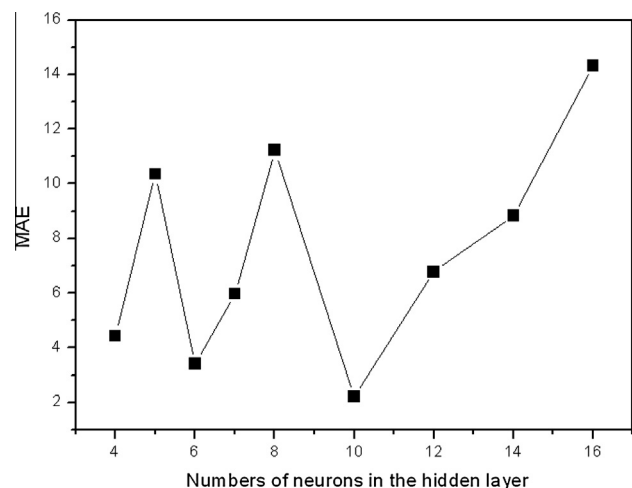


Fig. 3. Neuron number effect in the hidden layer on the MAE curve.

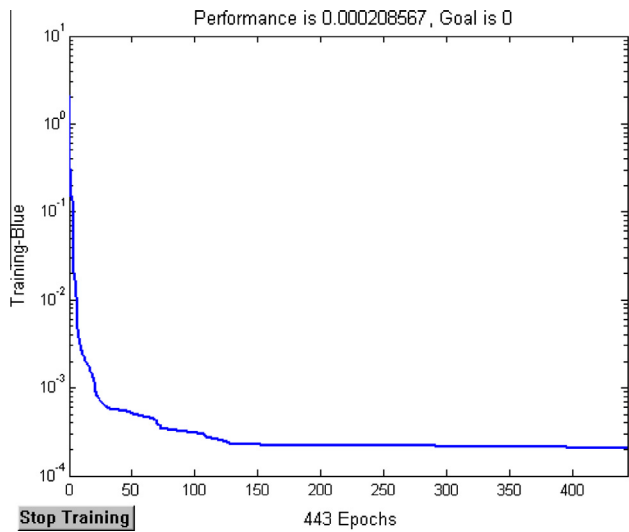


Fig. 4. Training convergence of the neural network model.

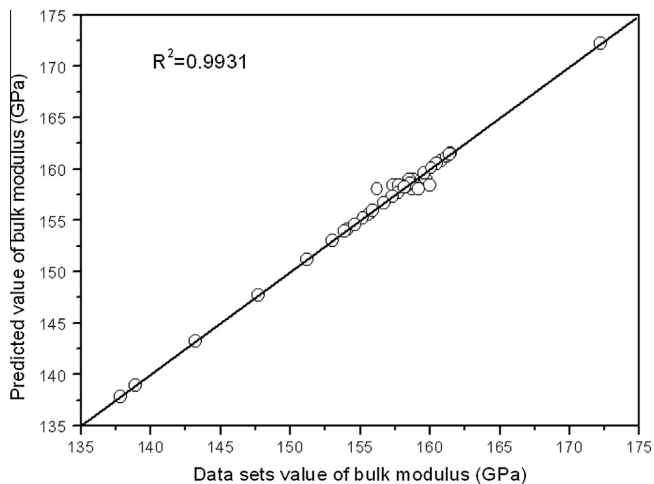


Fig. 5. ANN model simulation results compared with datasets values.

3.2. GA-ANN result

The optimization problem is based on a GA, with a single point crossover. Each individual was constructed with fitness function which obtained from the trained ANN model. In order to implement the genetic algorithm, some tests were performed to represent the combination of different options of crossover probability (P_c), mutation probability (P_m), selection operator, population size and generation number. Following the current investigation, the initial population was selected to be 50 and the P_c and P_m were chosen to be 0.8 and 0.001 respectively. The roulette-wheel and tournament for the selection operators were used. The GA runs were carried out to evaluate the concentration of manganese for optimizing the bulk modulus. Using the mentioned flowchart of GA-ANN model, presented in Fig. 2, the chemical composition of Cr and Ni are fixed for amount of about (18.50 wt%) each. Only the Mn concentration varies, the manganese alloying element is constructed optimization problem, which is conducted by GA. The evolution of the algorithm with successive generations has been shown in Fig. 6. Starting from the fitness value of about (24.06%), the representative curve has been increased quickly for the first 15th generation, reached the level of about 25% at the 25th generation and then it still invariant. Through the mentioned

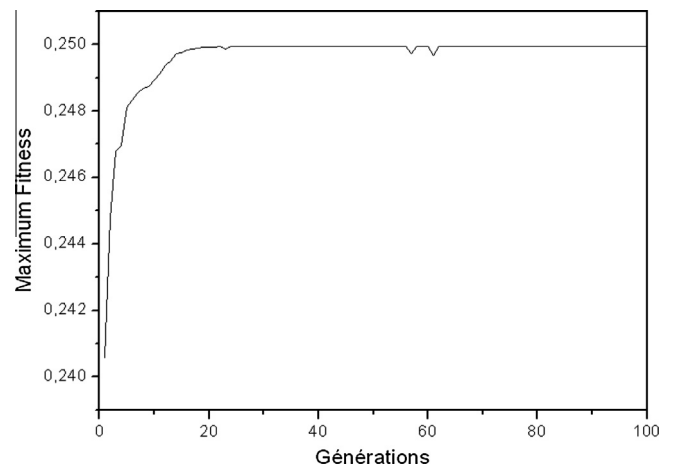


Fig. 6. Variation of fitness function with generations.

Table 2

Bulk modulus (B) and Mn concentration through the GA-ANN and DFT.

	X = Mn (wt%)	B (GPa)
$\text{Fe}_{0.5}\text{Cr}_{0.185}\text{Ni}_{0.185}\text{Mn}_x$		
GA-ANN approach	12.46	163.60
DFT calculation	12.50	159.00

curve, it can be noted that the GA has reached a maximum value after the 100th generation. The acquired results are used to optimize the concentration of Mn in FeCrNiMn alloy. From the above fixed values of Cr and Ni, The obtained concentration of manganese in weight percentage is about 12.46 (wt%). This indicates that when the concentration of Mn is 12.46% in the $(\text{Fe}(\text{bal})\text{Cr}_{0.185}\text{Ni}_{0.185}\text{Mn}_{0.1246})$ alloy, the predicting value of bulk modulus is 163.60 GPa.

3.3. Comparison between GA-ANN and computational theory (DFT)

In order to evaluate the potential and ability of the GA-ANN algorithm, the validation by mean of the computational method based on DFT was carried out. The model was described by 16-atom cubic supercell built from $2 \times 2 \times 1$ fcc. First, a Fe supercell with 16 atoms was created [17] and then, a part of Fe atoms were randomly substituted by the Cr, Ni and Mn atoms to obtain an alloy, which is $\text{Fe}_8\text{Cr}_3\text{Ni}_3\text{Mn}_2$ ($\text{Fe}_{0.5}\text{Cr}_{0.185}\text{Ni}_{0.185}\text{Mn}_{0.125}$). The cut off energy was chosen as 420 eV and $9 \times 9 \times 9$ k -points meshes are used. Spin polarized was considered in the calculation. By mean of Eq. (4) a bulk modulus is determined by fitting using the four parameters Birch–Murnaghan equation of state. The bulk modulus B was found to be order of 159.00 GPa, listed in Table 2. According to the mentioned table, one can see the average relative error between the values determined through GA-ANN model and computational theory is within 2.89%.

4. Conclusion

The aim of this paper was to use the technique of data mining, based on the combined genetic algorithm (GA) and artificial neural network (ANN). An ANN model with three layers with 10 neurons in the hidden layer has been shown as an effective structure in modeling the complex relationship between the chemical composition and bulk modulus. It seems a powerful method for the bulk modulus prediction of the FeCrNiMn austenitic stainless steel alloys. The optimization model with the combined method of GA

and ANN has been developed and used for optimizing the concentration of manganese (Mn). The obtained Mn concentration produces a bulk modulus (B) round 163.60 GPa. The last value has been checked by DFT. The relative error recorded is approximately 2.89%. Through the comparison of the obtained results, the combined method of GA and ANN methodology showed the capability to be a powerful and proficient means to solve many non-linear problems of material science. Moreover, this approach can be used as an alternative in other experimental and computational methods, thereby in reducing the running time. Although the GA-ANN approach gives satisfactory results in the case of mechanical properties, whereas, the DFT remains the most widely used approach in the field of material science.

Acknowledgement

Special words of thanks and acknowledgements are the best contributors to my colleague Azzedine Boutelhig, from the PV water pumping Laboratory/URAER, for reviewing together the article.

References

- [1] Itshree Mohanty, Debashish Bhattacharjee, Shubhabrata Datta, *Computational Materials Science* 50 (2011) 2331–2337.
- [2] Mst Kamrunnahar, Mirna Urquidi-Macdonald, *Corrosion Science* 52 (2010) 669–677.
- [3] David Corne, Clarisse Dhaenens, Laetitia Jourdan, *European Journal of Operational Research* 221 (2012) 469–479.
- [4] Kemal Tusat Yücel, Cengiz Özel, *Advances in Engineering Software* 45 (2012) 54–60.
- [5] Narinoba Kagami, Ryuichiro Iwamoto, Tetsuji Tani, *Fuel* 84 (2005) 279–285.
- [6] K. Usha Rani, *International Journal of Data Mining and Knowledge Management Process (IJDKP)* 1 (5) (2011).
- [7] M.V.J. Suresh, K.S. Reddy, Ajit Kumar Kolar, *Applied Energy* 88 (2011) 4867–4873.
- [8] R. Edwin Raj, B.S.S. Daniel, *Computational Materials Science* 43 (2008) 767–769.
- [9] Mehmet Sirac Ozerdem, Sedat Kolkusa, *Materials and Design* 30 (2009) 764–769.
- [10] Mohammed Hayajneh, Adel Mahmood Hassan, Abdalla Alrashdan, Ahmed Turki Mayyas, *Journal of Alloys and Compounds* 470 (2009) 584–588.
- [11] H. Hasan Örkücü, Hasan Bal, *Expert Systems with Applications* 38 (2011) 3703–3709.
- [12] Machavaram Rajendra, Prakash Chandra Jena, Hifjuz Raheman, *Fuel* 88 (2009) 868–875.
- [13] R.G. Song, Q.Z. Zhang, *Journal of Materials Processing Technology* 117 (2001) 84–88.
- [14] M. Yazdanmehr, S.H. Mousavi Anijdan, A. Bahrami, *Computational Materials Science* 44 (2009) 1218–1221.
- [15] Y. Sun, W.D. Zeng, Y.F. Han, X. Ma, Y.Q. Zhao, *Computational Materials Science* 50 (2011) 1064–1069.
- [16] S.H. Mousavi Anijdan, H.R. Madaah-Hosseini, A. Bahrami, *Materials and Design* 28 (2007) 609–615.
- [17] K. Benyelloul, H. Aourag, *Computational Materials Science* 67 (2013) 353–358.
- [18] G. Dini, A. Najafzadeh, S.M. Monir-Vaghefi, A. Ebnonnasir, *Computational Materials Science* 45 (2009) 959–965.
- [19] M. Yazdanmehr, S.H. Mousavi Anijdan, A. Bahrami, *Computational Materials Science* 44 (2009) 1218–1221.
- [20] Z. Dahoo, Z.M. Sbartai, A. Castel, F. Ghomari, *Engineering Structures* 31 (2009) 1724–1733.
- [21] M. Izadifar, M. Zolghadri Jahromi, *Journal of Food Engineering* 78 (2007) 1–8.
- [22] S.H. Mousavi Anijdan, A. Bahrami, H.R. Madaah Hosseini, A. Shafyeyi, *Materials and Design* 27 (2006) 605–609.
- [23] Li. Aijun, Li. Hejun, Li. Kezhi, Gu Zhengbing, *Acta Materialia* 52 (2004) 299–305.
- [24] S.F. Fang, M.P. Wang, W.H. Qi, F. Zheng, *Computational Materials Science* 44 (2008) 647–655.
- [25] Shouchun Wang, Xiucheng Dong, Renjin Sun, *Expert Systems with Applications* 37 (2010) 4768–4771.
- [26] A. Durand, O. Devos, C. Ruckebusch, J.P. Huvenne, *Analytica Chimica Acta* 595 (2007) 72–79.
- [27] M. Zakeri, A. Bahrami, S.H. Mousavi Anijdan, *Materials and Design* 28 (2007) 2034–2039.
- [28] James A. Freeman, David M. Skapura, *Neural Networks Algorithms, Application, and Programming Techniques*, Addison-Wesley Publishing Company, Inc., 1991.
- [29] Shen Changyu, Wang Lixia, Li Qian, *Journal of Materials Processing Technology* 183 (2007) 412–418.
- [30] G. Kresse, D. Joubert, *Physical Review B* 59 (1999) 1758–1775.
- [31] P.E. Blochl, *Physical Review B* 50 (24) (1994) 17953.
- [32] G. Kresse, J. Furthmuller, *Physical Review B* 54 (1996) 11169–11186.
- [33] G. Kresse, J. Furthmuller, *Computational Materials Science* 6 (1996) 15–50.
- [34] J.P. Perdew, Y. Wang, *Physics Review B* 45 (1992) 13244–13249.
- [35] J.P. Perdew, J.A. Chevary, S.H. Vosko, K.A. Jackson, M.R. Pederson, D.J. Singh, C. Fiollhais, *Physics Review B* 46 (11) (1992) 6671.
- [36] P.O. Adebayo, K.M. Bamgbose, J.A. Olowofela, J.A. Oguntuase, G.A. Adebayo, *Physica B* 405 (2010) 4578–4581.
- [37] Shun-Li shang, Yi Wang, D.E. Kim, Zi-Kui Liu, *Computational Materials Science* 47 (2010) 1040–1048.
- [38] H.M. Ledbetter, *Journal of Materials Science* 20 (1985) 2923–2929.
- [39] H.M. Ledbetter, N.V. Frederick, M.W. Austin, *Journal of Applied Physics* 51 (1980) 305–309.
- [40] Sudook Kim, Hassel Ledbetter, *Journal of Materials Science* 29 (1994) 5462–5466.
- [41] H.M. Ledbetter, M.W. Austin, *Materials Science and Engineering* 70 (1985) 143–149.
- [42] S. Lin, H. Ledbetter, *Materials Science and Engineering A167* (1993) 81–85.
- [43] A. Bahrami, S.H. Mousavi Anijdan, H.R. Madaah Hosseini, A. Shafyeyi, R. Narimani, *Computational Materials Science* 34 (2005) 335–341.
- [44] S.H. Mousavi Anijdan, A. Bahrami, *Materials Science and Engineering A* 396 (2005) 138–142.

Résumé

La théorie de la fonctionnelle de la densité (DFT), implémentée dans le code de calcul Vienna Ab-initio Simulation Package (VASP). A été utilisée afin de calculer de l'énergie de l'état fondamental et les propriétés élastiques de certains alliages d'aciers inoxydables objet de notre travail. L'étude de l'alliage ternaire dont la composition est $Fe_{0.62}Cr_{0.185}Ni_{0.185}$, nous a permis de déterminer la constante du réseau correspondant à l'état fondamentale, le module de compressibilité, le module de compressibilité. Suite à cela les trois constantes élastiques C_{ij} ont été calculées par la méthode de la conservation du volume de la maille sous l'effet de déformations (en anglais : *volume conserving distortions methods*). En deuxième tentative nous avons appliqué les différentes méthodes du datamining à savoir les réseaux de neurones et les algorithmes génétiques, il s'agit d'une part d'évaluer la capacité de ces approches à effectuer les tâches de simulations et de prévisions des différentes propriétés élastiques des matériaux étudiés ; et d'autre part de comparer leurs résultats obtenus avec ceux donnés dans la littérature, et la méthode DFT présenté au chapitre précédent.

Mots clés : calcul ab initio, réseau de neurones, algorithme génétique, constantes élastiques

Abstract

In this paper, two methods were applied to determine the different elastic constants of the face centred cubic austenitic stainless steel $Fe_{0.62}Cr_{0.185}Ni_{0.185}$. Firstly, the quantum mechanical simulation was applied based on the first principles calculations within the generalized gradient approximation (GGA) by using the efficient strain-stress method. Secondly an artificial neural network (ANN) is used based on back propagation algorithm training. ANN model has been developed for the analysis and simulation of the correlation between the elastic properties and composition. In the training model three input layers each accept the weight percentage of the alloy component (Fe, Cr and Ni), while the three different elastic constants C_{ij} were employed as outputs. Different models of ANN were developed to predict the elastic constants. The performance indices such as coefficient of determination, mean square error were used to control the performance of the prediction capacity of the models developed in this study. In addition to this, elastic constants obtained from ANN models were compared with those obtained from quantum mechanical simulation and with those reported in the literature. The prediction results obtained by the two methods seem to be satisfactory.

Key words: first principles calculation, artificial neural network, genetic algorithm, elastic constants.

ملخص

في هذه الورقة، تم تطبيق طريقتين لتحديد الأشرطة المطاطية مختلفة الثوابت من الوجه تركزت مكعب الأوستنيتي الفولاذ المقاوم للصدأ $Fe_{0.62}Cr_{0.185}Ni_{0.185}$. أولاً، تم تطبيق محاكاة ميكانيكية الكم استناداً إلى المبادئ الأولى الحسابات داخل تقريب الانحدار المعمم (GGA) باستخدام طريقة فعالة سلاله من الإجهاد. ثانياً شبكة العصبية الاصطناعية (ANN) يستخدم يعتمد على ظهره التدريب خوارزمية الانتشار. وقد تم تطوير نموذج ANN لتحليل ومحاكاة للعلاقة بين خصائص المرونة وتكوينها. في نموذج التدريب ثلاث طبقات المدخلات كل قبول نسبة الوزن للعنصر سبيكة (الحديد، والكروم والنيكل)، بينما كانوا يعملون على ثلاث ثوابت المطاطية المختلفة، وذلك كنتائج. وقد وضعت نماذج مختلفة من ANN للتنبؤ الثوابت المرنة. مؤشرات الأداء مثل معامل التحديد، يعني استخدمت مربع الخطأ للسيطرة على أداء لقدرة التنبؤ من النماذج المتقدمة في هذه الدراسة. وبالإضافة إلى ذلك، تمت مقارنة الثوابت المرنة التي تم الحصول عليها من نماذج ANN مع تلك التي تم الحصول عليها من الكم محاكاة الميكانيكية ومع تلك التي ذكرت في الأدبيات. نتائج التنبؤ التي حصل عليها طريقتين يبدو أن تكون مرضية.

الكلمات الدالة : حساب المبادئ الأولية، الشبكة العصبية الاصطناعية، الخوارزمية الجينية، الثوابت المرنة