

Université de Montréal

Utilisation de l'apprentissage automatique pour
approximer l'énergie d'échange-corrélation

par

Étienne Cuierrier

Département de chimie
Faculté des arts et des sciences

Thèse présentée en vue de l'obtention du grade de
Philosophiæ Doctor (Ph.D.)
en Chimie

12 Janvier 2024

© Étienne Cuierrier, 2024

Université de Montréal

Faculté des arts et des sciences

Cette thèse intitulée

Utilisation de l'apprentissage automatique pour approximer l'énergie d'échange-corrélation

présentée par

Étienne Cuierrier

a été évaluée par un jury composé des personnes suivantes :

Radu Iftimie

(président-rapporteur)

Matthias Ernzerhof

(directeur de recherche)

Richard Martel

(membre du jury)

Hilke Bahmann

(examinatrice externe)

Radu Iftimie

(représentant du doyen de la FESP)

Résumé

Le sujet de cette thèse est le développement de nouvelles approximations à l'énergie d'échange-corrélation (XC) en théorie de la fonctionnelle de la densité (DFT).

La DFT calcule l'énergie électronique d'une molécule à partir de la densité électronique, une quantité qui est considérablement plus simple que la fonction d'onde. Cette théorie a été développée durant les années 1960 et elle est devenue la méthode de choix en chimie quantique depuis 1990, grâce à un ratio coût/précision très favorable. En pratique, la DFT est utilisée par les chercheurs et l'industrie pour prédire des spectres infrarouges, la longueur des liens chimiques, les barrières d'activation, etc. Selon l'approche Kohn-Sham, seulement le terme de l'énergie XC est inconnu et doit être approximé. Les chapitres de ce texte sont des articles consacrés au développement d'approches non locales et à l'utilisation de l'apprentissage automatique pour améliorer la précision et/ou la rapidité des calculs de l'énergie XC.

Le premier article de cette thèse concerne le développement d'approximations non locales au trou XC [Cuierrier, Roy, et Ernzerhof, JCP (2021)]. Notre groupe de recherche a précédemment développé la méthode du facteur de corrélation (CFX) [Pavlíková Přecechtělová, Bahmann, Kaupp, et Ernzerhof, JCP (2015)] et malgré les résultats supérieurs de CFX comparativement aux fonctionnelles courantes en DFT pour le calcul de l'énergie, cette approche n'est pas exacte pour les systèmes uniélectroniques. Les méthodes non locales telles que le facteur X [Antaya, Zhou, et Ernzerhof, PRA (2014)] corrigent ce problème. Ainsi, le but du projet du premier article est de combiner CFX avec le facteur X, afin de former un facteur XC exact pour l'atome d'hydrogène, tout en conservant les bonnes prédictions de CFX pour les molécules. Nos résultats montrent que notre modèle non local est exact pour les systèmes uniélectroniques, cependant, la densité électronique non locale a un comportement fortement oscillatoire qui rend difficile la construction du facteur XC et la qualité de ses prédictions pour les propriétés moléculaires est inférieure aux fonctionnelles hybrides. Notre étude permet de fournir une explication concernant l'échec des méthodes non locales en chimie, comme l'approximation de la densité pondérée [Gunnarsson, Jonson, et Lundqvist, PLA (1976)]. Les nombreuses oscillations de la densité non locale limitent la performance des facteurs XC qui sont trop simples et qui ne peuvent pas atténuer ces oscillations.

Le sujet du deuxième article de cette thèse [Cuierrier, Roy et Ernzerhof, JCP (2021)] est relié aux difficultés rencontrées durant le premier projet. L'apprentissage automatique (ML) est devenu une méthode populaire dans tous les domaines de la science. Les réseaux de neurones artificiels (NN) sont particulièrement puissants, puisqu'ils permettent un contrôle et une flexibilité considérables lors de la construction de fonctions approximatives. Ainsi, nous utilisons un NN pour modéliser le trou X à partir de contraintes physiques. Durant le premier projet de cette thèse, nous avons observé qu'une fonction mathématique simple n'est pas adaptée pour être combinée avec la densité non locale, les NN pourraient donc être un outil utile pour approximer un trou X. Néanmoins, ce chapitre s'intéresse à la densité locale, avant de s'attaquer à la non-localité. Les résultats que nous avons obtenus lors du calcul des énergies X des atomes montrent le potentiel des NN pour construire automatiquement des modèles du trou X. Une deuxième partie de l'article suggère qu'un NN permet d'ajouter d'autres contraintes à des approximations du trou X déjà existantes, ce qui serait utile pour améliorer CFX. Sans les NN, il est difficile de trouver une équation analytique pour accomplir cette tâche. L'utilisation du ML est encore récente en DFT, mais ce projet a contribué à montrer que les NN ont beaucoup d'avenir dans le domaine de la construction de trou XC.

Finalement, le dernier chapitre concerne un projet [Cuierrier, Roy, Wang, et Ernzerhof, JCP (2022)] qui utilise aussi des NN en DFT. Des travaux précédents du groupe ont montré que le terme de quatrième ordre du développement en série de puissances de la distance interélectronique du trou X ($T_\sigma(\mathbf{r})$) [Wang, Zhou, et Ernzerhof, PRA (2017)] est un ingrédient utile pour améliorer les approximations du calcul de l'énergie X pour les molécules. Cependant, il n'a pas été possible de construire un modèle qui satisfait le deuxième et le quatrième terme du développement en série de puissances simultanément. Ainsi, avec l'expertise développée en apprentissage automatique lors du deuxième projet, le but de l'étude du troisième article est d'utiliser $T_\sigma(\mathbf{r})$ comme une variable d'entrée à un NN qui approxime l'énergie X. Nous avons montré qu'en utilisant comme ingrédients la fonctionnelle de PBE, $T_\sigma(\mathbf{r})$ et un NN, il est possible de s'approcher de la qualité des résultats d'une fonctionnelle hybride (PBEh) pour le calcul d'énergies d'atomisation, de barrières d'activation et de prédiction de la densité électronique. Cette étude démontre que $T_\sigma(\mathbf{r})$ contient de l'information utile pour le développement de nouvelles fonctionnelles en DFT. $T_\sigma(\mathbf{r})$ est en principe plus rapide à calculer que l'échange exact, donc nos fonctionnelles pourraient s'approcher de l'exactitude d'une fonctionnelle hybride, tout en étant plus rapides à calculer.

Mots clés: Théorie de la fonctionnelle de la densité, énergie et trou d'échange-corrélation, apprentissage automatique, réseaux de neurones artificiels.

Abstract

The subject of this thesis is the development of new approximations for the exchange-correlation (XC) energy in Density Functional Theory (DFT).

DFT calculates the electronic energy from the electronic density, which is a considerably simpler quantity than the wave function. DFT was developed during the 1960s and became the most popular method in quantum chemistry during the 1990s, thanks to its favourable cost/precision ratio. In practice, DFT is used by scientists and the industry to predict infrared spectra, bond lengths, activation energies, etc. The Kohn-Sham approach in DFT is by far the most popular, since it is exact in theory and only the XC functional has to be approximated. The exact form of the XC functional is unknown, thus the development of new approximations for the XC functional is an important field of theoretical chemistry. In this thesis, we will describe the development of new non-local methods and the use of machine learning to improve the prediction and the efficiency of the calculation of XC energy.

The first article in this thesis [Cuierrier, Roy, and Ernzerhof, JCP (2021)] is about the development of non-local approximations of the XC hole. Our research group previously developed the correlation factor approach (CFX) [Pavlíková Přecechtělová, Bahmann, Kaupp, and Ernzerhof, JCP (2015)]. The prediction of CFX for molecular properties compares favourably to other common functionals. However, CFX suffers from one-electron self-interaction error (SIE). Non-local models such as the X factor [Antaya, Zhou, and Ernzerhof, PRA (2014)] can fix the SIE, thus the goal of this project is to combine CFX with the X factor to build a non-local XC factor. We show that our method is exact for one-electron systems, however, our simple XC factor is not appropriate for the oscillatory behaviour of the non-local density and the results for molecules are inferior when compared to hybrid functionals. Our study provides an explanation of why non-local models, such as the weighted density approximation [Gunnarsson, Jonson, and Lundqvist, PLA (1976)], are not as successful as the common DFT functionals (PBE, B3LYP, etc.) in chemistry. The non-local electronic density is an elaborate function and often has a large number of local minima and maxima. The development of functionals using simple XC factors does not lead to satisfying results for the prediction of molecular energies. Therefore, a sophisticated XC factor that could attenuate the oscillatory shape of the non-local density is required.

The second article [Cuierrier, Roy, and Ernzerhof, JCP(2021)] addresses the difficulties observed for the development of non-local functionals during the first project. Machine learning (ML) is a useful technique that is gaining popularity in many fields of science, including DFT. Neural networks (NN) are particularly powerful, since their structure allows considerable flexibility to approximate functions. Thus, in this chapter, we use a NN to approximate the X hole by considering many of its known physical and mathematical constraints during the training of the NN. The results we obtain, for the calculation of energies of atoms using the NN, reveal the potential of this method for the automation of the construction of X holes. The second part of the paper shows that an NN can be used to add more constraints to an existing X hole approximation, which would be quite useful to improve CFX. The X hole obtained for a stretched H₂ molecule is promising when compared to the exact values. ML is still a new tool in DFT and our work shows that it has considerable potential for the construction of XC hole approximations.

Finally, the last chapter [Cuierrier, Roy, Wang, and Ernzerhof, JCP (2022)] describes a project that also uses NN. In a previous work by our group, it is shown that the fourth-order term of the expansion of the X hole ($T_\sigma(\mathbf{r})$) could improve the calculation of the X energy for molecules [Wang, Zhou, and Ernzerhof, PRA (2017)]. However, developing an equation that satisfies both the second and fourth-order terms simultaneously proved difficult. Thus, using the expertise in ML we developed during the second project, we build a new NN that uses the fourth-order term of the expansion of the X hole as a new ingredient to approximate the XC energy. Starting from the PBE functional, we trained a NN to reproduce the X energy of the hybrid functional PBEh. Our results show that this approach is a considerable improvement compared to PBE for the calculation of atomization energies, barrier heights and the prediction of electronic density. This study confirms that the fourth-order term of the expansion of the X hole does include useful information to build functionals in DFT. Since the calculation of the fourth-order term has a more favourable computational scaling compared to the exact exchange energy, our new functionals could lead to faster calculations in DFT.

Keywords: Density functional theory, exchange-correlation energy, exchange-correlation hole, machine learning, artificial neural networks.

Table des matières

Résumé	v
Abstract	vii
Liste des tableaux	xiii
Liste des figures	xvii
Liste des sigles et des abréviations	xxi
Remerciements	xxvii
Introduction	1
1. Calcul quantique	1
1.1. Théorie de la fonction d'onde	2
1.1.1. <i>Principe variationnel</i>	4
1.1.2. <i>Théorie Hartree-Fock</i>	5
1.1.3. <i>Méthodes post-Hartree-Fock</i>	9
1.1.4. <i>Champ multiconfigurationnel autocohérent</i>	10
1.1.5. <i>Méthode du cluster couplé</i>	13
1.2. Théorie de la fonctionnelle de la densité	14
1.2.1. <i>Théorèmes de Hohenberg-Kohn</i>	16
1.2.2. <i>Recherche sous contrainte de Levy-Lieb</i>	17
1.2.3. <i>Approche Kohn-Sham</i>	18
1.3. Connexion adiabatique	21
1.4. Trou d'échange et de corrélation	21
1.4.1. <i>Propriétés du trou d'échange-corrélation</i>	24
1.5. Approximations de l'énergie d'échange-corrélation	28
1.5.1. <i>Les fonctionnelles et le spin</i>	31
1.5.2. <i>Approximation de la densité locale</i>	31
1.5.3. <i>Approximation du gradient généralisé</i>	33
1.5.4. <i>Méta-approximation du gradient généralisé</i>	34

1.5.5.	<i>Fonctionelles hybrides</i>	36
1.5.6.	<i>Fonctionelles hybrides locales</i>	39
1.5.7.	<i>Fonctionelles hybrides par séparation de portée</i>	40
1.5.8.	<i>Approximation de la densité pondérée</i>	42
1.5.9.	<i>Problèmes courants des approximations à l'énergie d'échange-corrélation</i>	44
1.6.	Ensembles de bases	47
2.	Apprentissage automatique	49
2.1.	<i>Apprentissage supervisé</i>	51
2.2.	<i>Apprentissage non supervisé</i>	53
2.3.	<i>Apprentissage par renforcement</i>	54
2.4.	Réseaux de neurones	55
2.4.1.	<i>Algorithme d'optimisation des paramètres d'un réseau de neurones</i>	57
2.5.	Apprentissage automatique appliqué à la théorie de la fonctionnelle de la densité	59
2.5.1.	<i>Notre philosophie concernant l'utilisation de l'apprentissage automatique pour approximer l'énergie d'échange-corrélation</i>	61
3.	Objectifs de la thèse	63
	Références bibliographiques	64
	Premier article. The factorization ansatz for non-local approximations to the exchange-correlation hole	75
1.	Introduction	77
2.	Non-local approximations of the exchange hole	79
3.	Lower and upper bound to the exchange energy per particle in terms of the non-local electron density	82
4.	Modeling static correlation within the non-local approach	83
5.	Modeling dynamical correlation	85
5.1.	The three parameter $f_C^3(\mathbf{r}, u)$ correlation factor	85
5.2.	Reviewing the CFX correlation factor	86
5.3.	Avoiding double-counting of correlation effects	87
6.	Computational details	87
7.	Assessing the non-local functionals	87

7.1. Accounting for exchange and static correlation	88
7.2. Dynamic correlation with correlation factors	91
7.3. Static correlation from MCSCF	94
8. Conclusions	99
9. Data Availability Statement	99
10. Appendix	100
10.1. Computation of the on-top value, curvature, and the energy density of the MCSCF XS hole.	100
References	102
Deuxième article. Constructing and representing exchange–correlation holes through artificial neural networks	
1. Introduction	107
2. Theory for generating exchange hole models from physical constraints	108
3. Examples of X hole models represented through neural networks	109
3.1. Simplest X hole model employing the on-top and normalization condition ...	110
3.2. ExMachina: Neural-network-generated X holes depending on the curvature..	111
3.3. Evolving a given X or XC hole through machine generated factors	118
4. Technical Details	120
4.1. Neural networks	120
4.2. Evaluation of spherically-averaged exchange holes	122
5. Conclusions	122
Acknowledgments	123
Author declarations	123
Conflict of Interest	123
Data Availability Statement	123
References	124
Troisième article. The fourth-order expansion of the exchange hole and neural networks to construct exchange-correlation functionals .	
1. Introduction	128

2.	The fourth-order expansion of the exchange hole	130
3.	Machine learning to construct functionals drawing on the fourth-order expansion of the exchange hole	131
3.1.	Alternative enhancement factor satisfying the homogeneous gas limit	134
4.	Assessing and analyzing NNx4 and related functionals	134
4.1.	Exchange contribution to atomization energies	134
4.2.	Atomization energies and barrier heights	135
4.3.	Self-consistent electronic densities	136
4.4.	Enhancement factor of a neural network functional	137
5.	Conclusions	140
6.	Supplementary Material	141
	Acknowledgments	141
7.	Data Availability Statement	141
S1.	Supplementary materials	142
S1.1.	Technical details of the neural network and the electronic structure calculations	142
S1.2.	The mean absolute errors of atomic energies	142
S1.3.	The fourth-order derivative of the exchange hole for the N ₂ molecule	142
S1.4.	Impact of the choice of activation function in the neural network	144
S1.5.	Basis set dependence of \mathcal{T}_σ -dependent functionals	146
S1.6.	Dissociation of H ₂ +	146
S1.7.	Self-consistent implementation of the NN functionals	146
S1.7.1.	Derivative of s with respect to P_{uv}	148
S1.7.2.	Derivative of Q with respect to P_{uv}	148
S1.7.3.	Derivative of T with respect to P_{uv}	149
	References	151
	Conclusions	155

Liste des tableaux

I-1	Résumé de plusieurs propriétés mathématiques des trous d'échange-corrélation, d'échange et de corrélation.	24
II-1	Exchange and exchange-correlation factors developed in this work. f_X^4 and f_X^5 represent the basic exchange factors, where f_X^5 contains one additional empirical parameter (λ) compared to f_X^4 . Both of these models are described in Section 2. To account for static correlation (XS denotes exchange plus static correlation), the opposite-spin static correlation (OS) and Multi-Configurational reference (MC) approaches are described in Section 4. Dynamic correlation is accounted for by two correlation factors (f_C^3 and f_C^{CFX}) which are detailed in Section 5.1 and 5.2, respectively. The various exchange and correlation factors are then combined as indicated in the table and applied to $\rho(\mathbf{r},u)$ to obtain XC functionals.	88
II-2	Total energies for atoms in Hartree units, MAE abbreviates Mean Absolute Error. The approximations considered are characterized by their factors f as described in Table II-1. For $f_{\text{XS}}^{5\text{OS}}$, $\lambda = 0.82$	89
II-3	Atomization energies in kcal/mol for a set of single and multi-bonded molecules. MAE and ME are the mean absolute error and mean error, respectively. For $f_{\text{XS}}^{5\text{OS}}$: $\lambda = 0.82$	90
II-4	Total energies for atoms in Hartree and the Mean Absolute Error (MAE) for different methods. For $f_C^{\text{CFX}} f_{\text{XS}}^{\text{OS}}$, $a = 0.13$ and $\lambda = 1.30$. For $f_C^3 f_{\text{XS}}^{\text{OS}}$, $\lambda = 0.82$	93
II-5	Atomization energies in kcal/mol for a set of single and multi-bonded molecules. MAE and ME are the mean absolute error and mean error, respectively. For $f_C^{\text{CFX}} f_{\text{XS}}^{5\text{OS}}$, $a = 0.13$ and $\lambda = 1.30$. For $f_C^3 f_{\text{XS}}^{5\text{OS}}$, $\lambda = 0.82$	94
II-6	Total energies for atoms in Hartree and the Mean Absolute Error(MAE) for different methods. For $f_C^{\text{CFX}} f_{\text{XS}}^{4\text{MC}}$, $a = 0.18$ the same as [21].	96
II-7	Atomization energies in kcal/mol for a set of single and multi-bonded molecules. MAE and ME are the mean absolute error and mean error, respectively. For $f_C^{\text{CFX}} f_{\text{XS}}^{4\text{MC}}$, $a = 0.18$ the same as [21].	97

III-1	Exchange energies for atoms in Hartree units. All approximations have been implemented self-consistently with respect to variations of the orbitals. ExMc and ExMc α are the functionals generated by the machine learning algorithm (ExMachina) proposed here. LDA denotes the local approximation derived from the homogeneous electron gas and BR represents the Becke-Roussel approximation. "Exact" stands for self-consistent Hartree-Fock exchange. MAE abbreviates mean absolute error.	117
IV-1	Errors in the X contributions to atomization energies (ΔE_X) relative to PBEh, i.e. $\Delta E_X^{approx.} - \Delta E_X^{PBEh}$. The mean absolute errors are calculated for the 140 molecules of the W4-11 data set [56, 57] and are listed in kcal/mol. The functionals are evaluated with self-consistent PBE orbitals.	135
IV-2	Calculated atomization energies in kcal/mol for the validation dataset of molecules with single and multiple bonds. MAE and ME are the mean average error and mean error, respectively. The calculations are done self-consistently for all functionals. The reference results are obtained from [60–62].	137
IV-3	Mean absolute errors (MAE) and maximal absolute error (MAX) in kcal/mol for an atomization energy dataset of 140 molecules (W4-11) and a barrier heights dataset of 76 systems (BH76). All the calculations are performed with self-consistent PBE orbitals.	138
IV-4	Errors in the self-consistent electron density relative to CCSD calculations (see Eq 4.1) for the dataset of molecules of Tab. IV-2.	138
IV.S1	The mean absolute errors (MAEs) of total energies of atoms ranging from H to Ar. For the models using neural networks, self-consistent PBE orbitals are used, while PBEh is calculated self-consistently. The reference energy are obtained from Refs.[70–72].	143
IV.S2	Errors in the X contributions to atomization energies relative to PBEh, i.e. $\Delta E_X^{approx.} - \Delta E_X^{PBEh}$. The mean absolute errors are calculated over the 140 molecules of the W4-11 data set [57] and is listed in kcal/mol. The calculation are done using self-consistent PBE orbitals.	144
IV.S3	Mean absolute errors (MAE) and maximal absolute error (MAX) in kcal/mol for the atomization energy dataset of 140 molecules (W4-11) and a barrier heights dataset of 76 systems (BH76). All the calculation are post-PBE.	144

IV.S4 Error on the calculated electron density compared to CCSD calculations (see equation 4.1 in the main text) for the dataset of molecules of Tab. IV-2 of the main text.....	145
IV.S5 The mean absolute errors (MAE) for total energies (E^{tot}) of atoms (H to Ar) and for the atomization energies (AE) of the 4 dataset [56, 57], obtained with different gaussian basis sets (def2-SVP [66], def2-TZVP [67], and cc-PVTZ [74–78]).....	146

Liste des figures

1.1	Schéma qui montre un système fermé traité selon la méthode RHF et de deux systèmes avec une couche ouverte qui sont représentés par ROHF et UHF respectivement.	8
1.2	Les courbes de dissociation de la molécule de H_2 pour la théorie RHF, UHF et CASSCF (l'espace actif choisi est [2,2]) avec l'ensemble de bases cc-PVTZ (voir la section 1.6 sur les ensembles de bases). r_{HH} représente la distance entre les deux atomes d'hydrogène.	12
1.3	Schéma de la procédure pour calculer l'énergie d'une façon autocohérente selon la méthode Kohn-Sham.	20
1.4	Esquisses des trous d'échange (X), de corrélation (C) et total pour la molécule de H_2 à sa longueur de lien d'équilibre. La position de l'électron de référence est représentée par le point qui est près de la position de l'atome H_A	27
1.5	Esquisses des trous d'échange (X), de corrélation (C) et total pour la molécule de H_2 étirée. La position de l'électron de référence est près de l'atome H_A	28
1.6	Échelle de Jacob de la DFT [56]. Les approximations des niveaux supérieurs vont généralement inclure les ingrédients des niveaux inférieurs. LDA est l'approximation de la densité locale et GGA est l'approximation du gradient généralisé.	30
1.7	Schéma qui illustre la connexion adiabatique et le fonctionnement des fonctionnelles hybrides. λ est le paramètre de la force de couplage, E_x est l'énergie d'échange, E_c est l'énergie de corrélation, T_c est la contribution de la corrélation à l'énergie cinétique et $U_{xc}(\lambda)$ est l'énergie potentielle d'échange-corrélation qui dépend de λ . DFA signifie <i>Density Functional Approximation</i> , donc approximation de la fonctionnelle de la densité.	37
1.8	Courbes de dissociation pour la molécule de H_2^+ pour la méthode d'Hartree-Fock ainsi que la fonctionnelle B3LYP. L'ensemble de bases cc-PVTZ est utilisé.	45

1.9	Comparaison entre une orbitale de type Slater (STO), une orbitale de type gaussien (GTO), une somme de deux GTO (STO-2G) et une somme de trois GTO [122]. Les paramètres utilisés dans les GTO proviennent de la référence [122].	48
1.10	Le compromis entre le biais et la variance. Le biais est associé à l'erreur lors de l'apprentissage sur les données d'entraînements, tandis que la variance est la sensibilité du modèle aux fluctuations aléatoires dans un ensemble de données. Un modèle trop simple pour un problème mène à une erreur importante causée par un biais élevé, ce qui correspond à un sous-apprentissage. À l'opposé, un modèle trop complexe pour un problème pourrait souffrir de surapprentissage et avoir une variance élevée. Un modèle qui a surpris des tendances dans les données n'est donc pas généralisable, malgré une erreur faible lors de l'entraînement.	51
1.11	Les données représentent une fonction quadratique avec un bruit aléatoire de type gaussien. Le modèle 1 est une régression linéaire, tandis que le modèle 2 est un polynôme de degré 30. Les paramètres des modèles ont été obtenus à l'aide d'un ajustement par moindres carrés. Le modèle 1 souffre de sous-apprentissage tandis que le modèle 2 a une variance élevée, donc il a surpris les tendances dans les données.	52
1.12	À gauche, un exemple de classification à l'aide de l'apprentissage supervisé. Une méthode ML est entraînée à partir de données qui sont représentées par des ronds et des étoiles, pour ensuite classifier des entrées futures afin de savoir si elles sont un rond ou une étoile. À droite, un exemple d'apprentissage non supervisé pour partitionner des données. Pour l'apprentissage non supervisé, c'est au ML de séparer les données, l'information concernant si les entrées sont des étoiles ou des ronds n'est pas présente.	53
1.13	Schéma qui explique le principe de l'apprentissage par renforcement. Un agent fait une action pour interagir avec l'environnement. Ce dernier va retourner une récompense ou une punition. L'état représente la situation actuelle de l'agent dans l'environnement.	54
1.14	Exemple schématique d'un réseau de neurones artificiels avec deux variables d'entrées, deux couches cachées de six neurones chacune, ainsi qu'une variable de sortie. Les lignes dans ce schéma représentent les poids. Ce schéma a été construit à l'aide de l'outil de la référence [140].	56
2.1	Comparison of exchange and exchange-plus-static-correlation holes for the N ₂ molecule. One N atom is at the position (0.00,0.00,0.00) and the other at	

	(0.00,0.00,2.08) in Bohr. z is the position along the molecular axis. CFX(X) indicates that the X hole of the CFX method is represented, the other approximations are identified with their factor f as described in Table II-1.	91
2.2	Comparison of exchange factors for the N ₂ molecule. One N atom is at the position (0.00,0.00,0.00) and the other at (0.00,0.00,2.08) in Bohr. z is the position along the molecular axis.....	92
2.3	Comparison of XC holes for the N ₂ molecule. One N atom is at the position (0.00,0.00,0.00) and the other at (0.00,0.00,2.08) in Bohr. z is the position along the molecular axis.....	95
2.4	Approximated XC and XS holes for the N ₂ molecule. One N atom is at the position (0.00,0.00,0.00) and the other at (0.00,0.00,2.08) in Bohr. z is the position along the molecular axis.....	98
3.1	Neural networks (NNs) for the representation of X and XC hole models. The NN on the left represents a reduced X hole $J_X^c(y,c)$ that depends on the reduced curvature (c) and electron separation (y). To satisfy further conditions, including conditions that introduce correlation effects, a second NN representing a correlation and/or exchange factor $\mathbf{f}^{ce}(y,c,\mathcal{E})$ is introduced. It transforms $J_X^c(y,c)$ into a new reduced hole model $J^{ce}(y,c,\mathcal{E}) = \mathbf{f}^{ce}(y,c,\mathcal{E}) \times J_X^c(y,c)$. In this example, the factor depends on a reduced energy density per particle $\mathcal{E} = \epsilon/ \epsilon_X^{\text{LDA}} $, ($\epsilon_X^{\text{LDA}} = -\frac{3}{4} \left(\frac{3}{\pi}\right)^{4/3} \rho^{1/3}$). $J_X^c(y,c)$ does not need to be represented by a NN, it can also be a conventional hole model. The final X or XC hole model is then $\rho_{X(C)}(u, \mathbf{c}, \epsilon) = \rho \times J^{ce}(k_F u, \frac{c}{\rho k_F^2}, \frac{\epsilon}{ \epsilon_X^{\text{LDA}} })$, where $k_F = (3\pi^2\rho)^{1/3}$. Further details about the NNs are provided in Section 4.1.	113
3.2	Comparison of $J_X^{\text{ExMc}}(y,c)$ (full lines), $J_X^{\text{ExMc}\alpha}(y,c)$ (dotted lines), and the Becke-Roussel $J_X^{\text{BR}}(y,c)$ (dashed lines) for various values of the reduced curvature $c = -4,0,4$. The reduced holes are plotted where $\rho_X(u, \mathbf{c}) = \rho J_X(k_F u, \frac{c}{\rho k_F^2})$	114
3.3	Exchange energies per particle in reduced units, $\epsilon_X/ \epsilon_X^{\text{LDA}} = \mathcal{E}_X$ (full lines) and entropy (dashed lines) of various functionals defined in the text versus the reduced curvature c	116
3.4	The electron density (ρ) of stretched H ₂ (bond length $R = 7.56$ Bohr) along the molecular axis. A and B label the positions at which the X holes are compared. .	120
3.5	Comparison of the reduced X holes of $\rho_X^{\text{ExMc}}(u,c,\epsilon)$ and $\rho_X^{\text{BR}}(u,c)$ with the exact reduced X hole. The system considered is the H ₂ molecule (bond length $R = 7.56$	

	Bohr). In addition to the reduced holes, the integrands of the first and second moment are compared as well. The reference positions A (left column) and B (right column) are as described in Fig. 3.4.....	121
4.1	The neural network used to represent the enhancement factor $f^{NNx4}(\mathbb{S}_\sigma, \mathbb{Q}_\sigma, \mathbb{T}_\sigma)$. This enhancement factor is then multiplied with the X energy per particle of the PBE approximation (ϵ_X^{PBE}) to yield a new X functional (ϵ_X^{NNx4}). The neural network consists of an input layer with the three variables $\mathbb{S}_\sigma, \mathbb{Q}_\sigma$, and \mathbb{T}_σ ; then there are two hidden layers that are twelve neurons wide; finally, there is a single output neuron. The figure is drawn using the tool of reference [53].....	133
4.2	The exchange enhancement factor ($\epsilon_{X, \sigma}/\epsilon_{X, \sigma}^{HEG}$) for PBE, B88, NNx0, and NNx4.	139
4.3	The exchange enhancement factor ($\epsilon_{X, \sigma}/\epsilon_{X, \sigma}^{HEG}$) for PBE, B88, NNHx0, and NNHx4.	140
4.S1	The values of the fourth derivative of the exchange hole $\mathcal{T}_\sigma(\mathbf{r})$ for the N_2 molecule along the z axis. The first atom is at the origin, while the second is at (0,0,1.10) Angstrom. The orange dashed lines mark the transition from the core to the valence region.....	143
4.S2	The exchange enhancement factor for PBE and the NNTx4 model for various \mathbb{Q}_σ and \mathcal{T}_σ	145
4.S3	Self-consistent calculation of the H_2^+ dissociation curve with NNx4 compared to the exact HF, PBEh, and PBE. The curves have been shifted such that the minima coincide.....	147

Liste des sigles et des abréviations

AC	Connexion adiabatique, de l'anglais <i>Adiabatic Connection</i>
AIM	Atomes dans les molécules, de l'anglais <i>Atoms in Molecules</i>
BO	Born-Oppenheimer
BR	Becke-Roussel
C	Corrélation
CASSCF	Méthode du champ autocohérent de l'espace actif complet, de l'anglais <i>Complete Active Space Self-Consistent Field</i>
CCSD	Cluster couplé simple double, de l'anglais <i>Coupled Cluster Single-Double</i>
DFA	Approximation de la fonctionnelle de la densité, de l'anglais <i>Density Functional Approximation</i>
DFT	Théorie de la fonctionnelle de la densité, de l'anglais <i>Density Functional Theory</i>

FCI	Configuration d'interaction complète , de l'anglais <i>Full Configuration Interaction</i>
HSE	Heyd-Scuseria-Erzerhof
GEA	Approximation de l'expansion du gradient, de l'anglais <i>Gradient Expansion Approximation</i>
GGA	Approximation du gradient généralisé, de l'anglais <i>Generalized Gradient Approximation</i>
GTO	Orbitales de type gaussiennes, de l'anglais <i>Gaussian Type Orbitals</i>
HF	Hartree-Fock
HEG	Gaz d'électrons homogène, de l'anglais <i>Homogeneous Electron Gas</i>
HK	Hohenberg-Kohn
LC	Corrigée à longue portée, de l'anglais <i>Long range Corrected</i>
LCAO	Combinaison linéaire d'orbitales atomiques, de l'anglais <i>Linear Combination of Atomic orbitales</i>

LDA	Approximation de la densité locale, de l'anglais <i>Local Density Approximation</i>
LH	Hybride locale, de l'anglais <i>Local Hybrid</i>
LMF	Fonction de mélange locale, de l'anglais <i>Local Mixing Function</i>
LR	Longue portée, de l'anglais <i>Long Range</i>
LRSH	Hybride par séparation de portée locale, de l'anglais <i>Local Range Separated Hybrid</i>
LSD	Spin-densité locale, de l'anglais <i>Local Spin-Density</i>
MAE	Erreur absolue moyenne, de l'anglais <i>Mean Absolute Error</i>
MCSCF	Champ multiconfigurationnel autocohérent, de l'anglais <i>Multi-Configurational Self-Consistent Field</i>
Méta-GGA	Méta-approximation du gradient généralisé, de l'anglais <i>Meta-Generalized Gradient Approximation</i>
MLP	Perceptron multicouches, de l'anglais <i>Multi-Layers Perceptron</i>
ML	Apprentissage automatique, de l'anglais <i>Machine Learning</i>

NN	Réseaux de neurones, de l'anglais <i>Neural Networks</i>
PBE	Perdew-Burke-Ernzerhof
PBEh	Perdew-Burke-Ernzerhof hybride
PVTZ	Valence polarisée de triple zêta, de l'anglais <i>Polarized Valence Triple Zeta</i>
RKS	Kohn-Sham restreinte, de l'anglais <i>Restricted Kohn-Sham</i>
RHF	Hartree-Fock restreinte, de l'anglais <i>Restricted Hartree-Fock</i>
RMSProp	Racine carrée de l'erreur quadratique moyenne, de l'anglais <i>Root Mean Squared Propagation</i>
RSH	Hybride par séparation de portée, de l'anglais <i>Range Separated Hybrid</i>
SIE	Erreur d'auto-interaction, de l'anglais <i>Self-Interaction Error</i>
SR	Courte portée, de l'anglais <i>Short Range</i>
STO	Orbitales de type Slater, de l'anglais <i>Slater Type Orbitals</i>
TFD	Thomas-Fermi-Dirac

TPSS	Tao-Perdew-Scuseria-Staroverov
UKS	Kohn-Sham non restreinte, de l'anglais <i>Unrestricted Kohn-Sham</i>
UHF	Hartree-Fock non restreinte, de l'anglais <i>Unrestricted Hartree-Fock</i>
WDA	Approximation de la densité pondérée, de l'anglais <i>Weighted Density Approximation</i>
WFT	Théorie de la fonction d'onde, de l'anglais <i>Wavefunction Theory</i>
XC	Échange-Corrélation
XS	Échange-Corrélation Statique
X	Échange

Remerciements

Je voudrais remercier le professeur Matthias Ernzerhof pour son aide et sa disponibilité tout au long de mon doctorat. Je remercie aussi mes collègues du groupe: Yongxi Zhou, Emna Azek, Rodrigo Wang, Alexandre Giguère, Pierre-Olivier Roy, Tobias Henkes, Mohamed Loutis, Ariane Caron et Inès Mary. Sans le support de mon superviseur et du groupe, la complétion de cette thèse n'aurait pas été possible. Je suis reconnaissant envers mon comité de thèse, les professeurs Radu Iftimie, Christian Reber et Richard Martel pour leurs conseils durant la thèse. Je veux également remercier la professeure Hilke Bahmann d'avoir accepté d'examiner cette thèse. Je remercie aussi mes compagnons de jeux vidéo, Charles, David et Gaël. Je suis reconnaissant envers le Conseil de recherches en sciences naturelles et en génie du Canada et le Fonds de recherche du Québec - nature et technologies pour des bourses durant mon doctorat. Finalement, un gros merci à ma famille pour leur patience et leur support durant toutes mes études.

Introduction

Ce qui est simple est toujours faux. Ce qui ne l'est pas est inutilisable.

Paul Valéry

Dans ce premier chapitre, les concepts théoriques nécessaires pour comprendre la thèse sont introduits. Le sujet de la première section de ce chapitre concerne les calculs quantiques en chimie computationnelle, tandis que la seconde section introduit les concepts de base de l'apprentissage automatique.

1. Calcul quantique

En chimie computationnelle quantique, il y a deux grandes familles de méthodes couramment utilisées pour prédire la structure électronique de la matière: la théorie de la fonction d'onde (WFT) et la théorie de la fonctionnelle de la densité (DFT). Ces méthodes permettent, par exemple, de calculer des barrières d'activation, simuler les orbitales moléculaires, prédire des spectres infrarouges, etc [1–3]. Dans cette section, nous allons commencer par décrire brièvement la WFT, tout particulièrement les méthodes comme celle d'Hartree-Fock (HF), du champ multiconfigurationnel autocohérent (MCSCF) et du cluster couplé (CC). Ensuite, la théorie de la fonctionnelle de la densité (DFT) sera introduite. Ce sous-chapitre va discuter des différentes approximations à l'énergie d'échange-corrélation (XC).

Pour plus de détails concernant la chimie computationnelle en général, les livres de Cramer [2] et de Jensen [3] sont pertinents. Le livre de Szabo & Ostlund est une bonne référence pour la WFT [4]. Des livres recommandés concernant la DFT sont ceux de Koch & Holthausen [1] et Parr & Yang [5]. Sauf un avis contraire, les unités atomiques seront utilisées dans cette thèse: $\hbar = m_e = e = 4\pi\epsilon_0 = 1$. La notation qui sera utilisée pour représenter les orbitales dans l'introduction de cette thèse est la même que dans le livre de Jansen [3], c'est-à-dire que χ représente une orbitale atomique (AO), ϕ est une orbitale moléculaire (MO), Φ est fonction d'onde d'un déterminant de Slater et le symbole Ψ est pour une fonction d'onde générale, par opposition au déterminant de Slater.

1.1. Théorie de la fonction d'onde

L'équation fondamentale de la WFT est celle de Schrödinger dépendante du temps

$$i \frac{\partial}{\partial t} \Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N, \mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_M, t) = \hat{H} \Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n, \mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_M, t) \quad (1.1)$$

où $\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N, \mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_M, t)$ est la fonction d'onde qui dépend des positions des N électrons (\mathbf{r}_i), des M noyaux (\mathbf{R}_K) et du temps t . \hat{H} est l'opérateur hamiltonien qui contient les termes pour obtenir l'énergie cinétique et potentielle d'un système. Cette équation décrit l'évolution dans le temps de la fonction d'onde Ψ , et quoique Ψ ne soit pas un observable physique, différents opérateurs appliqués à celle-ci permettent d'obtenir des valeurs physiques. Ψ est la quantité fondamentale en WFT et elle contient toute l'information nécessaire pour décrire un système chimique.

Pour les projets réalisés durant cette thèse, nous n'étudions pas l'évolution dans le temps de Ψ , ainsi, nous allons nous concentrer sur l'équation Shrödinger indépendante du temps

$$\hat{H} \Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N, \mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_M) = E \Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n, \mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_M) \quad (1.2)$$

où E est l'énergie totale du système. Pour les hamiltoniens qui ne dépendent pas de t , l'équation de Shrödinger peut être résolue par séparation de variables. On assume qu'on peut factoriser la fonction d'onde en une partie qui ne dépend que du temps et une partie qui ne dépend que des positions, puis on obtient l'équation 1.2.

L'opérateur \hat{H} donnant l'énergie est séparable en deux opérateurs, l'un pour l'énergie cinétique (\hat{T}) et l'autre pour celle potentielle (\hat{V})

$$\hat{H} = \hat{T} + \hat{V}. \quad (1.3)$$

En l'absence d'un champ électrique ou magnétique, les énergies cinétique et potentielle d'une molécule sont composées des termes suivants

$$\hat{H} = \underbrace{-\frac{1}{2} \sum_{i=1}^N \nabla_i^2}_{\text{Cinétique e}} - \underbrace{\frac{1}{2} \sum_{K=1}^M \frac{1}{M_K} \nabla_K^2}_{\text{Cinétique noyaux}} - \underbrace{\sum_{i=1}^N \sum_{K=1}^M \frac{Z_K}{r_{iK}}}_{\text{Potentiel e-noyaux}} + \underbrace{\sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}}}_{\text{Potentiel e-e}} + \underbrace{\sum_{K=1}^M \sum_{B>K}^M \frac{Z_K Z_B}{R_{KB}}}_{\text{Potentiel noyau-noyau}} \quad (1.4)$$

où la lettre e signifie électrons, r_{ab} (R_{AB} pour les noyaux) représente la distance entre les particules a et b , Z_K est la charge et M_K est la masse du noyau K . On remarque que l'équation précédente a des termes d'énergie associés aux noyaux et aux électrons. Il existe une approximation qui sépare le mouvement des électrons de celui des noyaux afin de simplifier le problème [6]. L'électron a une masse beaucoup plus faible que celle d'un noyau ($m_p/m_e \approx 1836$ où m_p est la masse d'un proton), donc le mouvement des électrons est beaucoup plus rapide que celui des noyaux. L'approximation de Born-Oppenheimer (BO) [6] considère que la réponse des électrons est instantanée par rapport au mouvement du noyau. En d'autres mots, le mouvement des électrons est tellement rapide par rapport au noyau,

qu'ils s'ajustent de façon immédiate suite au déplacement du noyau. Les coordonnées nucléaires (\mathbf{R}) sont alors séparées de celles des électrons (\mathbf{r}) ce qui simplifie l'hamiltonien (Eq. 1.4). Ainsi, nous pouvons écrire l'équation de Schrödinger électronique

$$\hat{H}_{el}\Psi_{el}(\mathbf{r}; \mathbf{R}) = E_{el}\Psi_{el}(\mathbf{r}; \mathbf{R}) \quad (1.5)$$

où E_{el} est l'énergie électronique. L'hamiltonien électronique est donné par

$$\hat{H}_{el} = \hat{T}_e(\mathbf{r}) + \hat{V}_{eN}(\mathbf{r}; \mathbf{R}) + \hat{V}_{ee}(\mathbf{r}) \quad (1.6)$$

où $\hat{T}_e(\mathbf{r})$ est l'opérateur qui décrit l'énergie cinétique des électrons, $\hat{V}_{eN}(\mathbf{r}; \mathbf{R})$ est celui pour l'attraction électron-noyau (le point-virgule qui sépare \mathbf{r} et \mathbf{R} dans l'équation signifie que les coordonnées nucléaires sont des paramètres et non des variables) et $\hat{V}_{ee}(\mathbf{r})$ est l'opérateur qui décrit les interactions électrons-électrons. En comparant l'équation 1.6 avec l'équation 1.4, on remarque que le deuxième et le dernier terme de la deuxième équation ne sont pas présents dans l'hamiltonien électronique. Donc, pour obtenir l'énergie totale avec la méthode BO, il faut additionner l'énergie électronique avec celle nucléaire

$$E_{tot} = E_{el} + E_{nuc}. \quad (1.7)$$

Les noyaux sont des charges fixes selon l'approximation BO, alors E_{nuc} est une fonction simple qui calcule la répulsion entre les noyaux sans qu'un couplage avec des électrons soit présent. Si la géométrie d'une molécule est fixe, E_{nuc} est une constante. L'approximation BO est aussi connue sous le nom d'approximation adiabatique et pour avoir les couplages électrons-noyaux, il faut alors des modèles appelés non adiabatiques [7]. Dans cette thèse, nous allons nous intéresser seulement aux systèmes pour lesquels l'approximation de BO est considérée comme étant valide. De plus, nous ne discuterons pas en profondeur des effets relativistes présents en chimie quantique, des effets qui sont particulièrement importants pour décrire la chimie des atomes lourds [8]. L'énergie cinétique des électrons de cœur des atomes lourds est très grande et la vitesse des électrons s'approche de celle de la lumière, d'où l'importance d'inclure la théorie de la relativité pour décrire ces atomes. Les modèles développés dans cette thèse sont appliqués aux atomes allant de l'hydrogène à l'argon, ainsi qu'à des petites molécules organiques simples, donc les effets relativistes peuvent être négligés.

La fonction d'onde Ψ montre plusieurs propriétés qui n'ont pas d'équivalents classiques. Les électrons sont des fermions avec un spin de 1/2, donc la fonction d'onde électronique a la caractéristique d'être antisymétrique lorsque deux électrons sont échangés (dorénavant, l'indice el dans les équations de la fonction d'onde et l'hamiltonien électronique est ignoré, l'approximation BO étant toujours utilisée)

$$\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_i, \dots, \mathbf{r}_j, \dots, \mathbf{r}_N) = -\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_j, \dots, \mathbf{r}_i, \dots, \mathbf{r}_N). \quad (1.8)$$

La fonction d'onde n'est pas un observable physique, cependant, son module au carré représente la probabilité que les électrons $1, 2, \dots, N$ se trouvent dans des éléments de volumes

$$|\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)|^2 d\mathbf{r}_1 d\mathbf{r}_2 \dots d\mathbf{r}_N. \quad (1.9)$$

Considérant que la fonction d'onde est reliée à la probabilité de trouver un ensemble d'électrons dans une configuration donnée, elle est normalisée à 1 pour simplifier les équations

$$1 = \int \dots \int |\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)|^2 d\mathbf{r}_1 d\mathbf{r}_2 \dots d\mathbf{r}_N \quad (1.10)$$

et selon la notation de Dirac, la normalisation est

$$\langle \Psi | \Psi \rangle = 1. \quad (1.11)$$

1.1.1. Principe variationnel.

L'équation de Schrödinger électronique est impossible à résoudre analytiquement pour des systèmes contenant plus d'un électron. Ce problème est connu sous le nom du problème à N corps en mathématique [9]. Il est donc nécessaire d'utiliser une approche systématique en chimie computationnelle, afin d'obtenir une énergie électronique optimale. Le principe variationnel [1–4] accomplit cette tâche et il est à la base des approximations de la WFT et de la DFT. Ce principe est plus facilement compris à l'aide de la notation de Dirac, où la valeur attendue de l'énergie est

$$E = \frac{\langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle}. \quad (1.12)$$

Le principe variationnel stipule que l'énergie obtenue à partir d'une fonction d'onde approximative sera plus grande ou égale à la valeur exacte de l'énergie de l'état fondamental [2, 3]

$$\frac{\langle \Psi_{\text{approx.}} | \hat{H} | \Psi_{\text{approx.}} \rangle}{\langle \Psi_{\text{approx.}} | \Psi_{\text{approx.}} \rangle} = E_{\text{approx.}} \geq E_0 = \frac{\langle \Psi_0 | \hat{H} | \Psi_0 \rangle}{\langle \Psi_0 | \Psi_0 \rangle}. \quad (1.13)$$

L'énergie $E_{\text{approx.}}$ est égale à E_0 si la fonction d'onde approximée est celle qui est exacte, donc $\Psi_{\text{approx.}} = \Psi_0$. Ainsi, en chimie quantique, la recette à suivre pour obtenir une énergie optimale, selon le principe variationnel, est de minimiser l'énergie à l'aide de différentes $\Psi_{\text{approx.}}$. Plus l'énergie est basse, plus $\Psi_{\text{approx.}}$ s'approche de celle qui est exacte. Pour représenter $\Psi_{\text{approx.}}$, il est alors possible d'utiliser une fonction ayant des coefficients qui sont optimisés en minimisant l'énergie. La fonction d'onde hypothétique doit tout de même respecter certaines contraintes physiques, par exemple, elle doit être continue sur son domaine et elle doit être intégrable. Un exemple pour approximer la fonction d'onde totale est d'utiliser une combinaison linéaire de fonctions d'onde

$$\Psi_{\text{approx}} = \sum_i c_i \psi_i \quad (1.14)$$

où les coefficients c_i sont optimisés pour minimiser l'énergie. L'optimisation s'effectue sous la contrainte que la fonction d'onde est normalisée, ce qui correspond à $\sum_i c_i^2 = 1$ si les ψ_i forment une base orthonormée.

1.1.2. Théorie Hartree-Fock.

La première méthode tirée de la WFT que nous allons discuter en détail est celle d'Hartree-Fock (HF) [10, 11]. La théorie HF considère le caractère fermionique des électrons, cependant la corrélation instantanée entre les électrons est ignorée. Bien que l'exactitude de la théorie HF est généralement trop petite pour la majorité des applications en chimie, cette théorie est essentielle à comprendre en chimie computationnelle.

HF est une théorie dite de champ moyen, où un électron va interagir avec un champ généré par tous les autres électrons. Son hamiltonien est alors composé d'opérateurs à un électron qui est nommé l'opérateur Fock (\hat{f}_i).

Les électrons sont non corrélés, ainsi intuitivement, nous pourrions penser que la fonction d'onde totale HF est la multiplication des fonctions d'onde unielectronique, mais ce n'est pourtant pas le cas en pratique

$$\Phi_{HF} \neq \phi_1 \phi_2 \dots \phi_N. \quad (1.15)$$

La fonction d'onde précédente n'est pas antisymétrique sous l'échange de deux électrons, elle n'est alors pas adaptée pour représenter des électrons. Pour respecter le caractère fermionique des électrons, le spin et les spin-orbitales sont introduits

$$\varphi_i(\mathbf{x}) = \phi_i(\mathbf{r})\sigma(s), \quad \sigma = \alpha, \beta \quad (1.16)$$

où $\phi(\mathbf{r})$ est une orbitale spatiale et $\sigma(s)$ représente la fonction de spin. Une façon ingénieuse de permettre à Φ_{HF} d'être antisymétrique est de représenter la fonction d'onde par un déterminant de Slater

$$\Phi_{HF} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \varphi_1(\mathbf{x}_1) & \dots & \varphi_N(\mathbf{x}_1) \\ \vdots & \ddots & \vdots \\ \varphi_1(\mathbf{x}_N) & \dots & \varphi_N(\mathbf{x}_N) \end{vmatrix}. \quad (1.17)$$

Un déterminant satisfait la condition de l'antisymétrie de la fonction d'onde (équation 1.8), car une permutation de lignes ou de colonnes (donc un échange de la position de deux électrons) change le signe du déterminant. Le but de la méthode HF est ainsi d'optimiser les spin-orbitales et leurs coefficients pour minimiser l'énergie électronique. Selon le principe variationnel (section 1.1.1), plus cette énergie est basse, plus elle s'approche de celle exacte. Un élément important à mentionner est le fait que HF est une théorie qui calcule l'énergie des molécules approximativement, donc l'énergie minimale obtenue, selon le principe variationnel, est plus grande ou égale à celle qui est exacte. La théorie HF est exacte seulement pour les systèmes unielectroniques, donc à l'exception de ceux-ci, $E_{HF} > E_0$. L'énergie HF

est donnée par l'équation suivante

$$\begin{aligned}
E_{HF} = & - \sum_i^N \int d\mathbf{x}_1 \varphi_i^*(\mathbf{x}_1) \frac{1}{2} \nabla^2 \varphi_i(\mathbf{x}_1) - \sum_i^N \int d\mathbf{x}_1 \varphi_i^*(\mathbf{x}_1) \sum_A^M \frac{Z_A}{r_{1A}} \varphi_i(\mathbf{x}_1) \\
& + \sum_{i,j} \frac{1}{2} \int \int d\mathbf{x}_1 d\mathbf{x}_2 \varphi_i^*(\mathbf{x}_1) \varphi_i(\mathbf{x}_1) \frac{1}{r_{12}} \varphi_j^*(\mathbf{x}_2) \varphi_j(\mathbf{x}_2) \\
& - \sum_{i,j} \frac{1}{2} \int \int d\mathbf{x}_1 d\mathbf{x}_2 \varphi_i^*(\mathbf{x}_1) \varphi_j(\mathbf{x}_1) \frac{1}{r_{12}} \varphi_j^*(\mathbf{x}_2) \varphi_i(\mathbf{x}_2) \quad (1.18)
\end{aligned}$$

où le premier terme correspond à l'énergie cinétique des électrons, le deuxième est pour l'attraction électrons-noyaux, le troisième est pour l'intégrale de Coulomb et le dernier pour l'intégrale d'échange. On remarque que l'intégrale d'échange donne une énergie négative, donc le caractère fermionique des électrons va diminuer la répulsion coulombienne entre les électrons, car ceux de mêmes spins vont s'éviter. De plus, il n'y a aucun terme qui décrit la corrélation de Coulomb entre les électrons. La répulsion coulombienne calculée par la théorie HF est surestimée, car les électrons qui sont corrélés ont davantage la tendance à s'éviter que ceux non interagissant.

Sous la forme d'un opérateur, l'équation HF est

$$\hat{f}(\mathbf{x}_1)_i \varphi_i(\mathbf{x}_1) = \epsilon_i^\sigma \varphi_i(\mathbf{x}_1), i = 1, 2, \dots, N. \quad (1.19)$$

où ϵ_i^σ sont les énergies des orbitales pour un spin σ et l'opérateur Fock est

$$\hat{f}_i(\mathbf{x}_1) = -\frac{1}{2} \nabla_i^2 - \sum_A^M \frac{Z_A}{r_{iA}} + \sum_j^N (\hat{J}_j(\mathbf{x}_1) - \hat{K}_j(\mathbf{x}_1)) \quad (1.20)$$

où le premier terme décrit l'énergie cinétique des électrons, le deuxième est pour l'énergie potentielle entre les électrons et le troisième est l'opérateur de Coulomb et le dernier est l'opérateur d'échange. L'opérateur de Coulomb est donné par

$$\hat{J}_j(\mathbf{x}_1) = \int d\mathbf{x}_2 |\varphi_j(\mathbf{x}_2)|^2 \frac{1}{r_{12}} \quad (1.21)$$

où il représente le potentiel de répulsion coulombienne ressenti par un électron à une position et spin \mathbf{x}_1 causé par un autre électron dans une spin-orbitale φ_j . Le deuxième terme (\hat{K}) est l'opérateur d'échange

$$\hat{K}_j(\mathbf{x}_1) \varphi_i(\mathbf{x}_1) = \int d\mathbf{x}_2 \varphi_j^*(\mathbf{x}_2) \frac{1}{r_{12}} \varphi_i(\mathbf{x}_2) \varphi_j(\mathbf{x}_1). \quad (1.22)$$

Cet opérateur se nomme opérateur d'échange, car il correspond au changement de variables entre deux spin-orbitales. Les électrons de mêmes spins ne peuvent pas être au même endroit dans l'espace, donc l'énergie d'échange n'est pas nulle uniquement pour les électrons de spins parallèles.

En pratique, pour pouvoir calculer l'énergie HF, la technique de la combinaison linéaire d'orbitales atomiques (LCAO) [1–4] est utilisée. La LCAO représente les orbitales moléculaires (MO) à partir d'une combinaison linéaire de bases atomiques telle que son nom le suggère. Une MO i est définie en termes de fonctions de base χ

$$\varphi_i(\mathbf{x}) = \sum_{\mu=1}^{N_{bases}} a_{\mu i}^{\sigma} \chi_{\mu}(\mathbf{r}) \quad (1.23)$$

où les $a_{\mu i}^{\sigma}$ sont des coefficients qui dépendent du spin σ et qui sont optimisés pour minimiser l'énergie. Selon le principe variationnel décrit à la section 1.1.1, l'énergie minimale d'un modèle sera égale ou supérieure à celle qui est exacte. Les coefficients dans la LCAO sont alors un paramètre pratique pour minimiser l'énergie. Lorsque l'opérateur $\hat{f}_i(\mathbf{x})$ est appliqué aux MO, on obtient

$$\hat{f}_i(\mathbf{x}_1) \sum_{\mu}^{N_{bases}} a_{\mu i}^{\sigma} \chi_{\mu}(\mathbf{r}) = \epsilon_i^{\sigma} \sum_{\mu}^{N_{bases}} a_{\mu i}^{\sigma} \chi_{\mu}(\mathbf{r}). \quad (1.24)$$

En multipliant l'équation précédente par la gauche par les fonctions de base et en intégrant, on obtient l'équation Pople-Nesbet [12] qui est sous une forme matricielle

$$\mathbf{F}^{\sigma} \mathbf{C}^{\sigma} = \mathbf{S} \mathbf{C}^{\sigma} \epsilon^{\sigma}, \quad (1.25)$$

où \mathbf{F}^{σ} est la matrice Fock pour un spin σ dont les éléments sont

$$F_{\mu\nu}^{\sigma} = \langle \chi_{\mu} | \mathbf{F}^{\sigma} | \chi_{\nu} \rangle, \quad (1.26)$$

\mathbf{S} est la matrice de recouvrement avec les éléments

$$S_{\mu\nu} = \langle \chi_{\mu} | \chi_{\nu} \rangle \quad (1.27)$$

et \mathbf{C}^{σ} est la matrice de coefficients. Pour déterminer les coefficients $a_{\mu i}^{\sigma}$, on commence par diagonaliser la matrice de recouvrement,

$$\mathbf{s} = \mathbf{U}^T \mathbf{S} \mathbf{U}, \quad (1.28)$$

où la matrice \mathbf{s} est diagonale. Ensuite, afin de retourner dans la base initiale de l'équation Pople-Nesbest, on a

$$\mathbf{S}^{-1/2} = \mathbf{U} \mathbf{s}^{-1/2} \mathbf{U}^{-1} \quad (1.29)$$

avec

$$\mathbf{S}^{-1/2} \mathbf{S} \mathbf{S}^{-1/2} = \mathbf{I}, \quad (1.30)$$

où \mathbf{I} est la matrice identité. On peut alors transformer la matrice Fock afin d'avoir une base orthonormale selon

$$\mathbf{F}'^{\sigma} = (\mathbf{S}^{-1/2})^T \mathbf{F}^{\sigma} \mathbf{S}^{-1/2}, \quad (1.31)$$

ainsi

$$\mathbf{F}'^{\sigma} \mathbf{C}'^{\sigma} = \epsilon^{\sigma} \mathbf{C}'^{\sigma}. \quad (1.32)$$

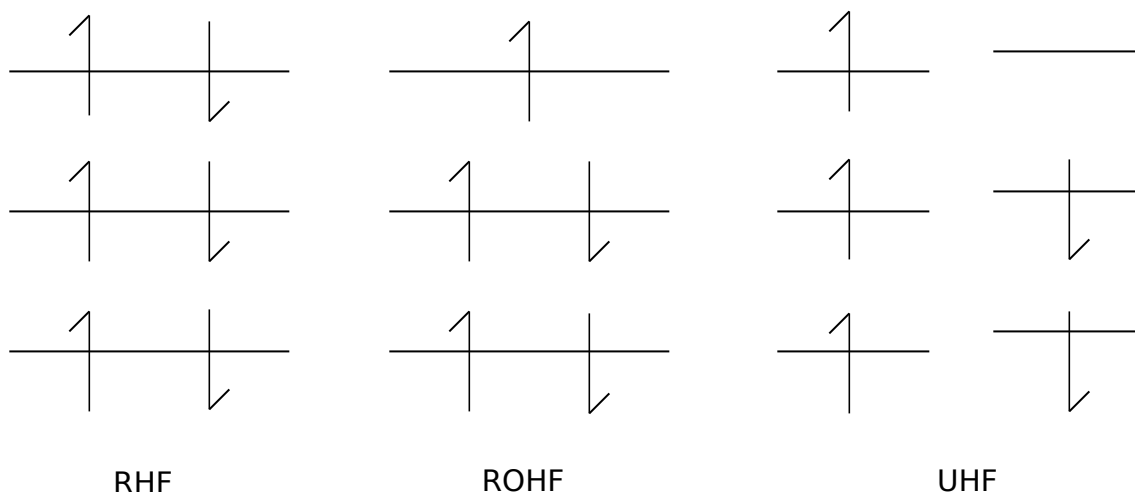


Fig. 1.1. Schéma qui montre un système fermé traité selon la méthode RHF et de deux systèmes avec une couche ouverte qui sont représentés par ROHF et UHF respectivement.

La diagonalisation de \mathbf{F}'^σ permet d'obtenir les énergies des orbitales (les valeurs propres) et leurs coefficients (les vecteurs propres), cependant, afin d'obtenir les coefficients dans la base originale (donc les orbitales canoniques) il faut appliquer la transformation suivante

$$\mathbf{C}^\sigma = \mathbf{S}^{-1/2} \mathbf{C}'^\sigma. \quad (1.33)$$

De plus, une observation importante est que la matrice Fock elle-même dépend des coefficients des orbitales spatiales. Il faut donc procéder par une méthode itérative qui se nomme le champ autocohérent pour minimiser l'énergie. Des paramètres initiaux sont définis pour les orbitales spatiales, puis l'opérateur Fock est appliqué pour calculer de nouveaux paramètres. Avec ces nouveaux paramètres, on réapplique l'opérateur $\hat{f}(\mathbf{x})$, etc. Ce processus se termine lorsque l'énergie a convergé à sa valeur minimale selon un critère de convergence défini par l'utilisateur.

En chimie computationnelle, les systèmes qui ont des électrons non appariés ne sont pas traités de la même façon que les molécules où les couches électroniques sont pleines [1–4] (voir la figure 1.1 pour un exemple visuel). Selon la méthode HF restreinte (RHF), les spin-orbitales de deux électrons appariés sont représentés par la même orbital spatiale $\phi(\mathbf{r}_1)$ et seulement le spin est différent entre les deux. Ainsi, l'équation de Pople-Nesbet (voir équation 1.25) est identique pour les spins α et β . Pour RHF, l'équivalent de l'équation Pople-Nesbet se nomme l'équation Roothan-Hall [13, 14].

Si le nombre d'électrons est impair et qu'il y a un électron non apparié, une possibilité est de considérer les électrons des couches pleines avec RHF, sauf pour le dernier électron. Cette méthode se nomme HF restreinte avec couche ouverte (ROHF).

La méthode la plus populaire pour décrire les molécules avec une couche électronique ouverte est que les spins α et β ne partagent pas la même orbitale spatiale (HF non restreinte

UHF), comme montré à la figure 1.1. Les orbitales pour les spins α et β sont ainsi séparées et calculées pour chaque spin individuellement. Dans les équations de cette section, nous avons utilisé le formalisme d'UHF. Cependant, un problème possible de la technique UHF se nomme la contamination de spin. La fonction d'onde UHF n'est pas nécessairement une fonction propre de l'opérateur \hat{S}^2 [2]. La contamination de spin signifie que la fonction d'onde est "contaminée" par des états de spins plus élevés. Par exemple, un état doublet pourrait avoir une composante de spin d'un état quartet non négligeable. Il est important de vérifier la valeur de $\langle \hat{S}^2 \rangle$ pour s'assurer que la valeur soit physiquement raisonnable.

Concernant les résultats obtenus par la méthode HF, pour la base de données nommée G2/97 [15] qui est constituée de valeurs d'énergies d'atomisation de 148 molécules organiques et inorganiques, l'erreur absolue moyenne (MAE) est de 159 kcal/mol [16] ce qui est une erreur très élevée. En chimie computationnelle, l'exactitude chimique est définie comme étant de 1 kcal/mol. Une fonctionnelle populaire de DFT comme B3LYP (voir la section 1.5.5) a une MAE de seulement 3 kcal/mol à titre de comparaison pour G2/97. Les énergies d'atomisation de HF sont considérablement sous-estimées dues à l'absence d'énergie de corrélation. HF est supérieure pour décrire les atomes comparativement aux molécules puisque l'énergie de corrélation électronique est relativement plus faible, donc, HF surestime davantage l'énergie électronique des molécules que des atomes.

Nous allons discuter dans la prochaine section (section 1.1.3) de la corrélation dynamique et statique. L'énergie de corrélation des atomes est majoritairement dynamique, alors que celle des molécules va aussi inclure de la corrélation statique.

1.1.3. Méthodes post-Hartree-Fock.

Les méthodes nommées post-Hartree-Fock (post-HF)[4] vont inclure l'énergie de corrélation électronique qui est définie comme étant l'énergie manquante à celle de HF

$$E_{corr.}^{WFT} = E_{exact} - E_{HF}. \quad (1.34)$$

$E_{corr.}^{WFT}$ est une valeur négative, car selon le principe variationnel, E_{HF} est plus grand (en d'autres mots moins négatif) que l'énergie électronique exacte. Même si $E_{corr.}^{WFT}$ ne représente qu'environ 1% de l'énergie totale, l'énergie de corrélation est essentielle pour correctement décrire des liens chimiques. Dans la section précédente, nous avons mentionné que les erreurs de la théorie HF en thermochimie sont trop grandes pour faire des prédictions fiables en chimie. La méthode de HF n'inclut pas les interactions instantanées entre les électrons, ces derniers ne subissent que la répulsion moyenne provenant des autres électrons. L'énergie de répulsion est alors surestimée, car des électrons corrélés vont avoir la tendance à s'éviter les uns les autres. Ce type de corrélation électronique se nomme la corrélation dynamique et elle est de courte portée. Un exemple typique est la corrélation qui se trouve entre les deux électrons de l'atome d'hélium.

Il y a un autre type de corrélation de longue portée qui se nomme la corrélation statique. La corrélation statique est moins intuitive à comprendre que celle dynamique. Elle survient lorsqu'un seul déterminant de Slater ne peut pas représenter la fonction d'onde correctement à cause de niveaux d'énergies dégénérés ou presque dégénérés. La corrélation statique explique pourquoi les métaux de transitions sont généralement des systèmes très difficiles à décrire en chimie quantique, puisqu'ils ont plusieurs orbitales de valence avec une énergie similaire [17]. Un autre exemple beaucoup plus simple à comprendre est la dissociation de molécules, comme la molécule de H_2 [1]. À l'équilibre, la corrélation est surtout de type dynamique et un seul déterminant de Slater est suffisant pour décrire correctement la molécule de H_2 . Les problèmes surviennent avec la théorie RHF à la limite de dissociation. Nous allons expliquer dans la section 1.1.4 qu'il y a une contribution ionique (alors $H^+ \cdots H^-$ et $H^- \cdots H^+$) à la fonction d'onde et celle-ci ne devrait pas être présente, puisqu'à la limite de dissociation, il faut obtenir deux atomes de H neutres. Avec deux déterminants de Slater, la contribution ionique est annulée et le système dissocié donne correctement deux atomes d'hydrogène. Une alternative à l'utilisation de plusieurs déterminants est de dissocier les molécules selon l'approche UHF. Il y a un atome avec un électron de spin α et l'autre de spin β à la limite de dissociation. UHF a cependant le défaut de briser la symétrie de la fonction d'onde en ayant des atomes de spins différents, quoique la limite de dissociation d'un point de vue énergétique est adéquate.

L'interaction de configuration complète (FCI) [3, 4] permet, en principe, de calculer l'énergie électronique de l'équation de Schrödinger indépendante du temps exactement, donc d'inclure autant la corrélation statique que celle dynamique. Elle consiste à utiliser une combinaison linéaire de déterminants de Slater obtenus en effectuant toutes les excitations possibles des électrons dans les orbitales virtuelles

$$\Psi_{FCI} = \sum_I c_I \Phi_I \quad (1.35)$$

où Φ_I sont les déterminants et c_I sont les coefficients à optimiser en minimisant l'énergie. Le nombre de déterminants croît factoriellement avec le nombre d'orbitales [3], donc FCI est utilisable seulement pour de très petits systèmes. Puisque son temps de calcul est trop grand en pratique, d'autres approches sont utilisées en WFT. Nous allons décrire deux de ces méthodes, la première est utilisée surtout pour représenter la corrélation statique (section 1.1.4), tandis que la deuxième (section 1.1.5) se concentre sur la corrélation dynamique. Il est néanmoins important de mentionner que la séparation de la corrélation en une composante statique et dynamique est souvent floue, donc ces deux méthodes vont tout de même inclure une partie de la corrélation dynamique ou statique [1].

1.1.4. *Champ multiconfigurationnel autocohérent.*

Comme mentionné dans la section précédente, l'approche HF n'inclut pas de corrélation

statique lorsque l'énergie est calculée, dû au fait que cette théorie n'utilise qu'un seul déterminant de Slater. Une méthode post-HF pour considérer la corrélation statique se nomme le champ multiconfigurationnel autocohérent (MCSCF) [2–4]. À l'instar de FCI, MCSCF utilise plusieurs déterminants pour construire la fonction d'onde,

$$\Psi_{MCSCF} = \sum_I^{N_{config.}} c_I \Phi_I \quad (1.36)$$

où Φ_I sont des déterminants de Slater, c_I sont des coefficients à déterminer et $N_{config.}$ signifie le nombre de configurations qui sont choisies par le chimiste. MCSCF est alors une technique qui demande une plus grande expertise à utiliser que HF, car c'est à l'utilisateur de déterminer les configurations qui sont représentées dans la fonction d'onde. Il faut donc inclure les configurations importantes pour correctement décrire la chimie du système. Comme pour FCI, le nombre de paramètres augmente très rapidement avec le nombre d'électrons et d'orbitales, il est donc nécessaire de réduire la taille de notre problème. L'approche du champ autocohérent dans l'espace actif complet (CASSCF) [2–4] est une méthode de MCSCF qui définit le choix des configurations. CASSCF consiste à diviser les MO dans un espace actif et un autre inactif. La notation utilisée pour décrire l'espace actif avec CASSCF est $[n, m]$, où n électrons sont distribués dans les m orbitales. Si le temps de calcul le permet, une approche généralement appropriée est d'inclure les électrons et les orbitales de valence. Mais le nombre de déterminants augmente de façon factorielle, donc en pratique, les calculs doivent avoir un nombre de configurations limité. De plus, une distinction avec FCI est que les coefficients des orbitales sont aussi optimisés.

Comme mentionné dans le paragraphe précédent, MCSCF ajoute de la corrélation statique dans les calculs comparativement à HF et le système le plus simple pour illustrer cette corrélation de longue portée est la molécule de H_2 étirée. Nous allons suivre la démarche du livre de Koch et Holthausen [1] et utiliser leurs notations. Selon la théorie HF, dans une base minimale, le déterminant de Slater indique que l'orbitale liante est occupée par deux électrons (le terme *det* désigne un déterminant)

$$\Phi_{HF} = \frac{1}{\sqrt{2}} \det\{\sigma_g(\mathbf{r}_1)\alpha(s_1)\sigma_g(\mathbf{r}_2)\beta(s_2)\} \quad (1.37)$$

où l'orbitale σ_g est représentable par une combinaison linéaire d'orbitales atomiques $1s$ des hydrogènes de gauche (L) et de droite (R)

$$\sigma_g = \frac{1}{\sqrt{2}}\{1s_L + 1s_R\} \quad (1.38)$$

et $\alpha(s_1)$ et $\beta(s_2)$ sont des fonctions de spins. Pour des électrons, nous avons

$$\alpha(1/2) = \beta(-1/2) = 1 \quad (1.39)$$

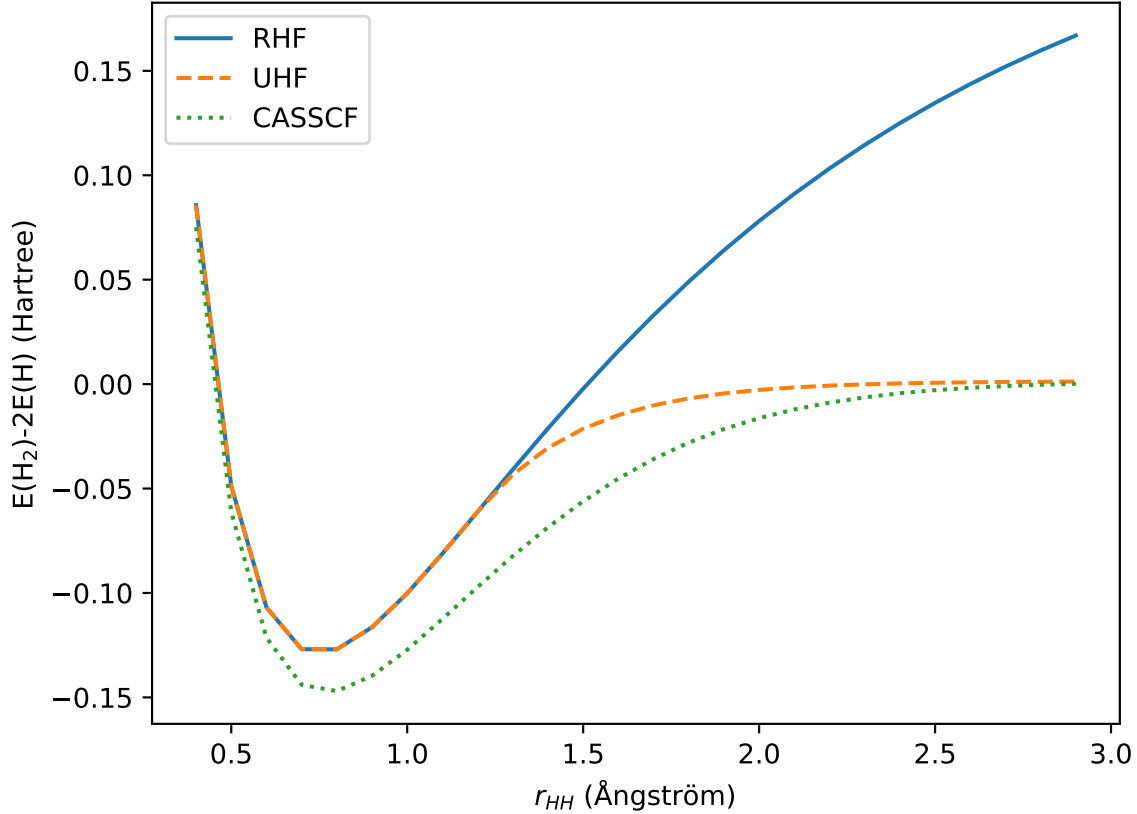


Fig. 1.2. Les courbes de dissociation de la molécule de H_2 pour la théorie RHF, UHF et CASSCF (l'espace actif choisi est $[2,2]$) avec l'ensemble de bases cc-PVTZ (voir la section 1.6 sur les ensembles de bases). r_{HH} représente la distance entre les deux atomes d'hydrogène.

et

$$\alpha(-1/2) = \beta(1/2) = 0. \quad (1.40)$$

En remplaçant σ_g du déterminant 1.37 par l'équation 1.38, on obtient l'équation

$$\Phi_{HF} = \frac{1}{2} [\det\{1s_L\alpha 1s_R\beta\} + \det\{1s_L\beta 1s_R\alpha\} + \det\{1s_L\alpha 1s_L\beta\} + \det\{1s_R\alpha 1s_R\beta\}] \quad (1.41)$$

où les deux premiers termes représentent une interaction covalente et les deux derniers sont des termes ioniques. La molécule de H_2 doit se dissocier en deux atomes de H neutre, les termes ioniques doivent être nuls en principe. Il n'est pas possible de les rendre nuls avec un seul déterminant de Slater sans briser la symétrie de la fonction d'onde (donc d'utiliser l'approche UHF). La figure 1.2 illustre l'échec de la théorie RHF pour dissocier le H_2 , la différence d'énergie avec celle de deux atomes neutres augmente et devient même positive plutôt que de tendre vers 0. Une solution alternative à l'utilisation d'UHF, afin de préserver la symétrie de la fonction d'onde, est de considérer plusieurs déterminants de Slater. Il est

donc possible d'ajouter l'orbitale antiliante à l'espace actif

$$\sigma_u = \frac{1}{\sqrt{2}}\{1s_L - 1s_R\} \quad (1.42)$$

ce qui va mener à cette fonction d'onde antisymétrique (AS)

$$\Phi_{AS} = \frac{1}{\sqrt{2}} \det\{\sigma_u(\mathbf{r}_1)\alpha(s_1)\sigma_u(\mathbf{r}_2)\beta(s_2)\}. \quad (1.43)$$

Ainsi, en additionnant le déterminant HF avec celui AS, la limite de dissociation que nous obtenons est

$$\Phi_{diss} = \frac{1}{\sqrt{2}}\{\Phi_{HF} + \Phi_{AS}\} = \frac{1}{2} [\det\{1s_L\alpha 1s_R\beta\} + \det\{1s_L\beta 1s_R\alpha\}], \quad (1.44)$$

donc, il n'y a pas de terme ionique. Comme montré à la figure 1.2, la corrélation statique est une interaction à longue portée. Une méthode comme CASSCF permet d'obtenir une courbe de dissociation comme celle d'UHF où l'énergie tend vers la valeur de zéro Hartree, tout en conservant la symétrie de la fonction d'onde.

1.1.5. Méthode du cluster couplé.

MCSCF est appropriée pour représenter la corrélation statique grâce à ces multiples déterminants de Slater. Cependant, pour pouvoir considérer la corrélation dynamique, la majorité des orbitales virtuelles doivent être tenues en compte. Une approche très populaire en WFT pour ajouter de la corrélation dynamique est la méthode du cluster couplé (CC) [2–4], en particulier, CC avec les excitations simples, doubles et avec les excitations triples qui sont calculées par la théorie de perturbation (CCSD(T)). CCSD(T) n'est pas la meilleure méthode en cas de corrélation statique très forte, mais cette méthode est considérée comme étant l'étalon-or en chimie computationnelle. La méthode CC ajoute de la corrélation entre les électrons à l'aide d'un opérateur cluster (\hat{T})

$$|\Psi_{CC}\rangle = e^{\hat{T}} |\Phi_{HF}\rangle = e^{\hat{T}} |\Phi_0\rangle \quad (1.45)$$

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \dots + \hat{T}_N \quad (1.46)$$

où N représente le nombre total d'électrons et \hat{T}_n est un opérateur qui génère les excitations de n électrons dans les orbitales virtuelles. Par exemple, lorsqu'il est appliqué à la fonction d'onde de référence, on obtient l'équation

$$\hat{T}_1 |\Phi_0\rangle = \sum_i \sum_a^{occ. virt.} t_i^a |\Phi_i^a\rangle \quad (1.47)$$

$$\hat{T}_2 |\Phi_0\rangle = \frac{1}{4} \sum_{i < j} \sum_{a < b}^{occ. virt.} t_{ij}^{ab} |\Phi_{ij}^{ab}\rangle \quad (1.48)$$

où les t_{ij}^{ab} sont des coefficients à déterminer, i, j sont des indices pour des orbitales occupées et a, b pour celles virtuelles et Φ est un déterminant de Slater. La fonction d'onde CC est

alors

$$|\Psi_{CC}\rangle = |\Phi_0\rangle + \sum_i \sum_a^{occ. virt} c_i^a |\Phi_i^a\rangle + \sum_{i<j} \sum_{a<b}^{occ. virt} c_{ij}^{ab} |\Phi_{ij}^{ab}\rangle + \dots \quad (1.49)$$

avec $c_i^a = t_i^a$ et $c_{ij}^{ab} = t_{ij}^{ab} + t_i^a t_j^b - t_j^a t_i^b$. Pour des raisons de temps de calcul, l'opérateur \hat{T} doit être tronqué. La méthode la plus utilisée est CCSD, qui inclut \hat{T}_1 (les excitations simples) et \hat{T}_2 (les excitations doubles), alors $\hat{T} \approx \hat{T}_1 + \hat{T}_2$. Le développement en série de Taylor de l'opérateur est

$$e^{\hat{T}_1 + \hat{T}_2} = \hat{1} + \hat{T}_1 + \left(\hat{T}_2 + \frac{1}{2}\hat{T}_1^2\right) + \left(\hat{T}_1\hat{T}_2 + \frac{1}{6}\hat{T}_1^3\right) + \dots \quad (1.50)$$

Pour trouver les coefficients de CCSD, la méthode variationnelle est appliquée avec les paramètres de l'opérateur cluster comme paramètres variationnels. Il est possible de montrer que l'énergie CCSD est finalement

$$E_{CCSD} = E_0 + \sum_{i<j} \sum_{a<b}^{occ. virt} c_{ij}^{ab} \langle \Phi_0 | \hat{H} | \Phi_{ij}^{ab} \rangle. \quad (1.51)$$

Le coût computationnel de CCSD est très élevé, avec une complexité de calcul qui croît en N^6 (N^4 pour HF à titre de comparaison) avec la taille de la molécule [4]. Il est aussi possible d'ajouter une correction à l'énergie, à l'aide d'une perturbation de troisième ordre, ce qui est nommé la théorie CCSD(T). CCSD(T) est considérée comme étant la méthode de choix en chimie quantique, pouvant atteindre des erreurs absolues moyennes de 1 kcal/mol pour le calcul d'énergies d'atomisation [18]. Bien entendu, le coût computationnel de CCSD(T) est élevé, donc les calculs sont effectués pour des molécules qui contiennent au plus une dizaine d'atomes. Mais comme mentionné précédemment, en cas de corrélation statique forte, MCSCF est une meilleure approche que CCSD(T), mais pour des systèmes à corrélation statique faible (par exemple, la plupart des molécules organiques à la géométrie d'équilibre), CCSD(T) est supérieure à MCSCF. Par exemple, pour la dissociation d'une chaîne de H_8 qui est un exemple de corrélation statique forte, la courbe de dissociation de CCSD(T) va diverger, contrairement à CASSCF dont la courbe converge correctement vers une constante [19].

1.2. Théorie de la fonctionnelle de la densité

La théorie de la fonctionnelle de la densité (DFT) [1, 16, 20, 21], le sujet principal de cette thèse, est la méthode la plus populaire en chimie quantique pour prédire les propriétés électroniques des molécules [16]. Plutôt que d'utiliser la fonction d'onde et de résoudre l'équation de Schrödinger, c'est la densité électronique ($\rho(\mathbf{r})$) qui est utilisée pour calculer l'énergie

$$E[\rho] = T[\rho] + V[\rho], \quad (1.52)$$

où $T[\rho]$ et $V[\rho]$ sont des fonctionnelles de la densité pour l'énergie cinétique et potentielle respectivement. Les crochets dans l'équation précédente représentent une fonctionnelle de la variable $\rho(\mathbf{r})$, c'est-à-dire, une fonction de fonction. La DFT est conceptuellement beaucoup plus simple que la WFT, puisque $\rho(\mathbf{r})$ ne dépend que des trois variables de position x, y, z , tandis que Ψ dépend de $4N$ variables ($3N$ variables pour les positions et N variables pour le spin) où N est le nombre d'électrons. De plus, Ψ n'est pas observable physiquement, contrairement à $\rho(\mathbf{r})$ qui peut être observée expérimentalement par des techniques de diffraction de rayon X [22].

La densité électronique montre plusieurs propriétés utiles qui permettent d'étudier les systèmes chimiques en profondeur [23]. Par exemple, $\rho(\mathbf{r})$ donne le nombre d'électrons lorsqu'on l'intègre dans l'espace

$$N = \int d\mathbf{r} \rho(\mathbf{r}) \quad (1.53)$$

et $\rho(\mathbf{r})$ permet de visuellement étudier la position des atomes dans une molécule. Le domaine de la chimie qui se nomme *Atoms in molecules* (AIM) [24] s'intéresse d'ailleurs à étudier la topologie de la densité électronique. AIM examine, par exemple, les liens chimiques dans une molécule à partir des caractéristiques de la topologie de la densité électronique dans cette région.

Également, plusieurs formules donnent les composantes de l'énergie potentielle à partir de $\rho(\mathbf{r})$. L'énergie d'attraction entre $\rho(\mathbf{r})$ et un noyau est donnée par l'équation

$$V_{ne}[\rho] = \sum_k^M \int d\mathbf{r} \frac{Z_k}{|\mathbf{r} - \mathbf{R}_k|} \rho(\mathbf{r}) \quad (1.54)$$

et l'énergie potentielle de répulsion entre deux densités de charge est

$$V_{ee}[\rho] = \frac{1}{2} \int \int d\mathbf{r}_1 d\mathbf{r}_2 \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{r_{12}}. \quad (1.55)$$

Concernant l'énergie cinétique, une équation a été trouvée pour un gaz d'électrons homogène (HEG) en 1927 par Thomas et Fermi (TF) [25, 26]. Les travaux de TF ont mené au premier modèle de DFT pour calculer l'énergie électronique

$$E_{TF} = \frac{3}{10} (3\pi^2)^{2/3} \int d\mathbf{r} \rho^{5/3}(\mathbf{r}) - Z \int d\mathbf{r} \frac{\rho(\mathbf{r})}{r} + \frac{1}{2} \int \int d\mathbf{r}_1 d\mathbf{r}_2 \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{r_{12}} \quad (1.56)$$

où le premier terme est l'énergie cinétique du HEG, le deuxième est l'attraction entre la densité et une charge uniforme Z et finalement, le dernier élément est la répulsion entre deux densités de charge. Nous avons alors un modèle qui calcule l'énergie électronique approximativement, uniquement à partir de $\rho(\mathbf{r})$.

Cependant, les énergies calculées à partir du modèle TF sont loin de bien représenter les molécules, l'approximation de l'énergie cinétique est beaucoup trop grossière. De plus, la corrélation électronique et l'énergie d'échange ne sont pas présentes. Néanmoins, Dirac en 1930 [27] a étendu la théorie pour ajouter l'énergie X (échange) du HEG et a obtenu la

formule suivante pour cette énergie

$$E_x^{HEG}[\rho] = -\frac{3}{4} (3\pi)^{1/3} \int d\mathbf{r} \rho^{4/3}(\mathbf{r}). \quad (1.57)$$

Le modèle Thomas-Fermi-Dirac (TFD) a eu peu d'impact en chimie, les méthodes provenant de la WFT donnent des résultats largement supérieurs à ceux de TFD [1]. Les molécules sont d'ailleurs instables selon le modèle de TFD. Il fallut attendre les années 1960 avant que des modèles DFT puissent rivaliser avec HF pour la modélisation moléculaire.

De plus, l'approche de TFD ne démontre pas qu'il est possible de remplacer la fonction d'onde par $\rho(\mathbf{r})$, ils ont seulement fait cette supposition qui semble plausible. La prochaine section va décrire le théorème qui a permis de prouver que $\rho(\mathbf{r})$ contient toute l'information nécessaire pour calculer l'énergie électronique, ce qui a contribué à l'obtention du prix Nobel de chimie en 1998 à Walter Kohn [28].

1.2.1. Théorèmes de Hohenberg-Kohn.

L'histoire de la DFT moderne débute en 1964 avec la publication des deux premiers théorèmes de Hohenberg-Kohn (HK) [21]. Le premier théorème montre que la densité électronique est une alternative valable à la fonction d'onde pour calculer l'énergie électronique, donc que $\rho(\mathbf{r})$ contient l'information nécessaire pour décrire les systèmes chimiques. Le théorème indique que le potentiel externe (V_{ext}) peut être totalement décrit par la densité $\rho(\mathbf{r})$ à l'état fondamental. Comme V_{ext} définit l'hamiltonien \hat{H} , $\rho(\mathbf{r})$ doit aussi permettre de définir entièrement l'état fondamental d'un système quantique. Pour prouver le premier théorème HK, le plus simple est de procéder par une preuve par l'absurde. On commence la preuve avec deux Hamiltoniens dont les potentiels externes diffèrent par plus qu'une constante additive, donc

$$\hat{H} = \hat{H} + \hat{V}_{ee} + \hat{V}_{ext} \quad (1.58)$$

et

$$\hat{H}' = \hat{H} + \hat{V}_{ee} + \hat{V}'_{ext}. \quad (1.59)$$

Ces deux Hamiltoniens vont correspondre à des fonctions d'onde distinctes Ψ et Ψ' , ainsi qu'à des valeurs d'énergies différentes E_0 et E'_0 . À la prochaine étape de la preuve, on suppose que ces deux hamiltoniens vont avoir la même densité électronique $\rho(\mathbf{r})$. Cette supposition permet d'observer une contradiction. Selon le principe variationnel, puisque Ψ et Ψ' sont différents, nous avons l'inéquation

$$E_0 < \frac{\langle \Psi' | \hat{H} | \Psi' \rangle}{\langle \Psi' | \Psi' \rangle} = \frac{\langle \Psi' | \hat{H}' | \Psi' \rangle}{\langle \Psi' | \Psi' \rangle} + \frac{\langle \Psi' | \hat{H} - \hat{H}' | \Psi' \rangle}{\langle \Psi' | \Psi' \rangle} \quad (1.60)$$

$$E_0 < E'_0 + \frac{\langle \Psi' | \hat{T} + \hat{V}_{ee} + \hat{V}_{ext} - \hat{T} - \hat{V}_{ee} - \hat{V}'_{ext} | \Psi' \rangle}{\langle \Psi' | \Psi' \rangle}, \quad (1.61)$$

ensuite, on peut aussi écrire la formule précédente comme

$$E_0 < E'_0 + \int d\mathbf{r} \rho(\mathbf{r})(v_{ext}(\mathbf{r}) - v'_{ext}(\mathbf{r})), \quad (1.62)$$

où il n'y a pas d'égalité, car les deux potentiels externes sont différents de plus d'une constante additive. Par symétrie, en changeant l'indice prime dans l'équation 1.62, on obtient l'équation

$$E'_0 < E_0 - \int d\mathbf{r} \rho(\mathbf{r})(v_{ext}(\mathbf{r}) - v'_{ext}(\mathbf{r})). \quad (1.63)$$

En additionnant les deux équations précédentes, la contradiction suivante est observée

$$E_0 + E'_0 < E'_0 + E_0. \quad (1.64)$$

Alors, nous avons montré que si V_{ext} est différent de V'_{ext} , la densité électronique de l'état fondamental ne peut pas être la même. En d'autres mots, il y a une correspondance directe entre la densité électronique et le potentiel externe. V_{ext} définit l'hamiltonien, ce qui détermine Ψ et ses propriétés comme l'énergie. Ce théorème est une preuve qu'un modèle comme TFD est plausible, puisque $\rho(\mathbf{r})$ contient l'information nécessaire pour calculer l'énergie.

Le deuxième théorème de HK concerne le principe variationnel pour la DFT. Le premier théorème a prouvé que la densité électronique peut remplacer la fonction d'onde, et maintenant nous allons prouver qu'il y a également un principe variationnel avec $\rho(\mathbf{r})$, comme il y a en un avec Ψ . Pour le prouver, il faut montrer que si $\rho(\mathbf{r})$ est différent de ρ_0 , nous avons

$$E[\rho_0] \leq E[\rho] = T[\rho] + E_{eN}[\rho] + E_{ee}[\rho] \quad (1.65)$$

et l'égalité est vraie lorsque $\rho = \rho_0$. Ayant déjà discuté du principe variationnel ainsi que le premier théorème de HK, la preuve est simple puisqu'il y a une correspondance entre la fonction d'onde et la densité. Nous obtenons finalement la formule

$$\frac{\langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle} = E[\rho] \geq E[\rho_0] = \frac{\langle \Psi_0 | \hat{H} | \Psi_0 \rangle}{\langle \Psi_0 | \Psi_0 \rangle}, \quad (1.66)$$

ce qui complète la description des deux théorèmes de HK. Ces preuves comportent cependant quelques défauts et restrictions qui doivent être corrigés. Selon l'approche de HK, la densité électronique doit être représentable par un V_{ext} , ce qui est une restriction assez compliquée à prouver. Il est très difficile de montrer que $\rho(\mathbf{r})$ est possiblement représentable par un certain V_{ext} d'un point de vue purement théorique, quoique cela semble raisonnable en pratique. Mel Levy [29] et Elliot Lieb [30] ont fait une preuve alternative à celle de HK qui est plus simple et généralisable.

1.2.2. Recherche sous contrainte de Levy-Lieb.

L'approche de Levy-Lieb [29, 30] décrit le principe variationnel en DFT de manière plus générale que celle de HK. Une minimisation est effectuée en cherchant sur toutes les Ψ qui permettent d'obtenir une densité, ensuite, une recherche est effectuée sur toutes les densités

qui sont N -représentables, donc qui intègrent au bon nombre d'électrons et qui sont des fonctions continues. De plus, cette double minimisation considère les états fondamentaux qui ont plusieurs densités possibles (donc des états dégénérés), ce qui n'est pas le cas avec l'approche de HK qui suppose un état non dégénéré. La formule de Levy-Lieb est la suivante (où l'on considère que Ψ est normalisée pour simplifier la formule)

$$E_0 = \min_{\rho \rightarrow N} \left(\min_{\Psi \rightarrow \rho} \langle \Psi | \hat{T} + \hat{V}_{ext} + \hat{V}_{ee} | \Psi \rangle \right), \quad (1.67)$$

on remarque la double minimisation qui est expliquée au paragraphe précédent. Le potentiel externe ne dépend que de $\rho(\mathbf{r})$, ainsi la minimisation sur les fonctions d'onde n'est pas nécessaire pour ce terme et on obtient

$$E_0 = \min_{\rho \rightarrow N} \left(\min_{\Psi \rightarrow \rho} \langle \Psi | \hat{T} + \hat{V}_{ee} | \Psi \rangle + \int d\mathbf{r} \rho(\mathbf{r}) v_{ext}(\mathbf{r}) \right). \quad (1.68)$$

Cette équation peut être simplifiée davantage grâce à l'introduction de la fonctionnelle universelle qui est définie comme

$$F_{uni.}[\rho] = \min_{\Psi \rightarrow \rho} \langle \Psi | \hat{T} + \hat{V}_{ee} | \Psi \rangle, \quad (1.69)$$

alors

$$E_0 = \min_{\rho \rightarrow N} \left(F_{uni.}[\rho] + \int d\mathbf{r} \rho(\mathbf{r}) v_{ext}(\mathbf{r}) \right). \quad (1.70)$$

L'équation précédente montre que contrairement à la méthode de HK, la densité n'a pas besoin d'être représentable par un potentiel externe (v représentable), alors il n'est pas nécessaire que $\rho(\mathbf{r})$ soit construite à partir d'un Ψ qui est une solution de l'équation de Schrödinger. Selon l'approche de Levy-Lieb, $\rho(\mathbf{r})$ doit seulement être représentable par un Ψ , ce qui est une contrainte beaucoup moins stricte que d'être v représentable.

1.2.3. Approche Kohn-Sham.

Les méthodes de HK et Levy-Lieb prouvent que $\rho(\mathbf{r})$ peut remplacer Ψ et qu'un principe variationnel existe en DFT, cependant, elles ne montrent pas comment procéder en pratique pour trouver ou approximer la fonctionnelle universelle $F_{uni}[\rho]$. De plus, il est évident qu'une double minimisation sur les fonctions d'onde, puis sur les densités, serait beaucoup trop longue pour pouvoir minimiser l'énergie électronique d'une molécule en pratique.

En 1965, Kohn et Sham (KS) [20] ont proposé une méthode théorique pour pouvoir calculer l'énergie électronique exacte, donc une approche alternative à la résolution de l'équation de Schrödinger. Cependant, la méthode KS réintroduit les orbitales pour trouver l'énergie cinétique d'électrons non corrélés (T_s)

$$T_s = - \sum_i \int d\mathbf{r} \phi_i^*(\mathbf{r}) \frac{1}{2} \nabla^2 \phi_i(\mathbf{r}). \quad (1.71)$$

Même si T_s ne contient pas la totalité de l'énergie cinétique, elle est une approximation largement supérieure à l'énergie cinétique comparativement à la méthode TFD [1]. La théorie

de KS montre plusieurs similarités avec celle de HF, où l'énergie cinétique est calculée avec des opérateurs à un électron. Kohn et Sham ont considéré un système fictif d'électrons non interagissant se déplaçant dans l'espace dans un potentiel modifié, ce dernier permettant de reproduire la densité électronique réelle. Les orbitales n'ont donc pas une signification physique comme pour la WFT, puisque les orbitales KS représentent un système fictif où les électrons sont non interagissant. Ces orbitales sont toute de même utiles pour calculer l'énergie cinétique sans corrélation électronique et reproduire la densité électronique réelle. Comme HF, la théorie KS-DFT va utiliser des opérateurs à un électron et un seul déterminant de Slater. Pourtant, KS-DFT est une théorie qui calcule l'énergie électronique exactement. Selon l'approche de KS, l'énergie électronique est donnée par la formule

$$E_{KS}[\rho] = T_s[\rho] + V_{eN}[\rho] + U[\rho] + E_{xc}[\rho] \quad (1.72)$$

où $T_s[\rho]$ est donné par l'équation 1.71, V_{eN} est l'énergie d'interaction entre les noyaux et les électrons, $U[\rho]$ est la répulsion coulombienne classique entre des électrons non corrélés et E_{xc} est l'énergie d'échange-corrélation (XC). Le dernier terme n'est qu'une fraction de l'énergie électronique totale, mais il est essentiel de bien le calculer pour décrire correctement les liaisons chimiques. Afin de calculer la densité électronique, la méthode KS utilise les orbitales

$$\rho(\mathbf{r}) = \sum_i^N |\phi_i(\mathbf{r})|^2. \quad (1.73)$$

On peut alors écrire la formule de l'énergie KS comme étant

$$E_{KS}[\rho] = - \sum_i \int d\mathbf{r} \phi_i^*(\mathbf{r}) \frac{1}{2} \nabla^2 \phi_i(\mathbf{r}) + \int d\mathbf{r} v_{ext}(\mathbf{r}) \rho(\mathbf{r}) + \int \int d\mathbf{r}_1 d\mathbf{r}_2 \frac{\rho(\mathbf{r}_1) \rho(\mathbf{r}_2)}{r_{12}} + E_{xc}[\rho] \quad (1.74)$$

Une définition de l'énergie XC est la somme des corrections associées aux énergies cinétique et potentielle

$$E_{xc}[\rho] = T_c[\rho] + U_{xc}[\rho]. \quad (1.75)$$

où $T_c[\rho] = T[\rho] - T_s[\rho]$ et $U_{xc}[\rho] = V_{ee}[\rho] - U[\rho]$ sont les corrections lorsque les interactions entre les électrons sont présentes pour l'énergie cinétique et l'énergie potentielle respectivement. Bien que l'approche KS-DFT soit exacte, le terme E_{xc} doit être approximé en pratique. L'énergie XC est séparable en une somme entre l'énergie d'échange et celle de corrélation selon l'équation

$$E_{xc}[\rho] = E_x[\rho] + E_c[\rho], \quad (1.76)$$

où l'énergie X exacte pour un déterminant de Slater est donnée par la formule

$$E_x = -\frac{1}{2} \sum_{\sigma} \sum_i \sum_j \int \int d\mathbf{r}_1 d\mathbf{r}_2 \phi_{i,\sigma}^*(\mathbf{r}_1) \phi_{j,\sigma}^*(\mathbf{r}_2) \frac{1}{r_{12}} \phi_{j,\sigma}(\mathbf{r}_1) \phi_{i,\sigma}(\mathbf{r}_2), \quad (1.77)$$

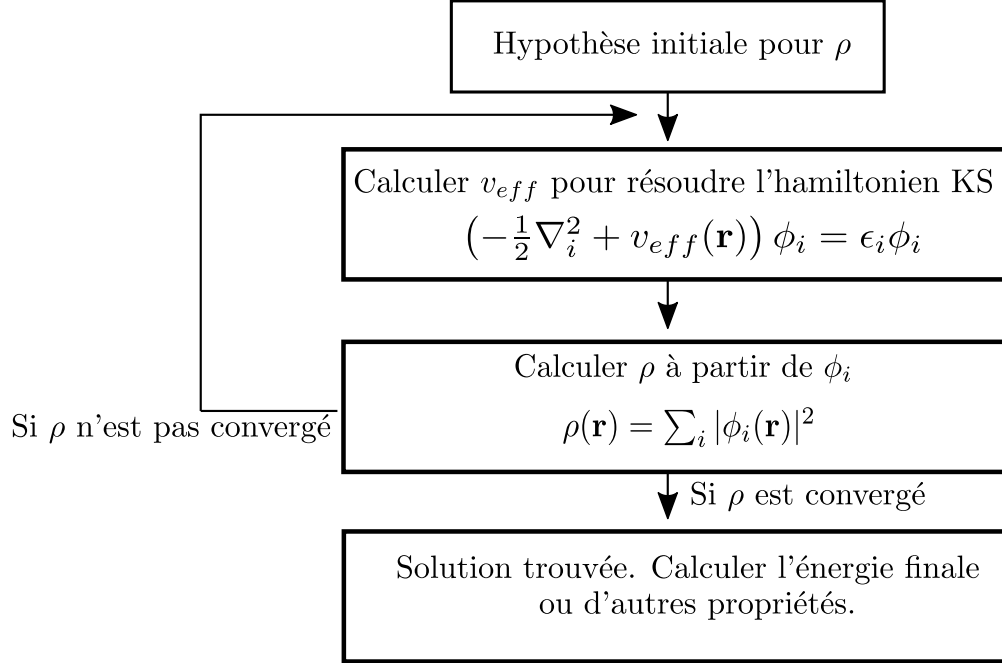


Fig. 1.3. Schéma de la procédure pour calculer l'énergie d'une façon autocohérente selon la méthode Kohn-Sham.

où l'échange est présent uniquement pour des orbitales d'un même spin σ . Une formule pour calculer l'énergie de corrélation selon la méthode KS sera donnée à la section 1.3. E_{xc} est généralement calculée à partir de l'équation

$$E_{xc}[\rho] = \int d\mathbf{r} \rho(\mathbf{r}) \epsilon_{xc}(\mathbf{r}) \quad (1.78)$$

où $\epsilon_{xc}(\mathbf{r})$ se nomme la densité d'énergie XC. Cette densité d'énergie par particule est multipliée par la densité à cette position. Un exemple d'approximation à l'énergie XC est celui de Dirac que nous avons discuté dans la section 1.2. Cette approximation n'inclut que l'énergie d'échange cependant.

En minimisant l'énergie de l'équation 1.74 par rapport aux orbitales, sous la contrainte que les orbitales sont normalisées, nous obtenons les équations KS

$$\left(\frac{-1}{2} \nabla_i^2 + v_{eff}(\mathbf{r}) \right) \phi_i(\mathbf{r}) = \epsilon_i \phi_i(\mathbf{r}) \quad (1.79)$$

$$v_{eff}(\mathbf{r}_1) = \int \frac{\rho(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_2 + v_{xc}(\mathbf{r}_1) - \sum_A^M \frac{Z_A}{r_{1A}}. \quad (1.80)$$

L'opérateur du potentiel effectif v_{eff} dépend de la densité, donc tout comme HF, un champ autocohérent est nécessaire pour minimiser l'énergie. La procédure autocohérente pour calculer l'énergie selon l'approche KS est montrée à la figure 1.3.

La méthode KS calcul l'énergie à partir de la densité électronique $\rho(\mathbf{r})$ totale, donc qui inclut des électrons des spins α et β . En l'absence de champ magnétique, il n'est pas

pas nécessaire d'avoir de l'information sur les spin-densités. La section 1.5.1 va discuter de la spin-DFT, une approche qui est essentielle en pratique pour obtenir des prédictions satisfaisantes des propriétés électroniques des molécules ayant des électrons non appariés.

1.3. Connexion adiabatique

Une composante difficile à approximer dans l'énergie XC est la contribution de la corrélation à l'énergie cinétique (T_c). Durant le milieu des années 1970, une méthode du nom de connexion adiabatique (AC) [31, 32] a été développée et elle a simplifié le calcul de l'énergie XC. Le fonctionnement de l'AC consiste à "connecter" le système Kohn-Sham non interagissant avec le système réel où les électrons interagissent selon un paramètre λ , qui va d'une valeur de 0 à 1 selon la force du couplage. L'hamiltonien est alors donné par la formule

$$\hat{H} = \hat{T} + \hat{V}_{ext}^\lambda + \lambda \hat{V}_{ee}, \quad (1.81)$$

où le potentiel externe \hat{V}_{ext}^λ dépend de la force du couplage λ , car il maintient la densité électronique équivalente à celle qui est exacte pour chaque λ : $\rho^\lambda(\mathbf{r}) = \rho_0(\mathbf{r})$. En examinant l'équation 1.81, on remarque les deux limites de λ : si $\lambda = 1$, on a l'hamiltonien d'un système d'électrons interagissant, mais lorsque $\lambda = 0$, les interactions entre les électrons sont nulles. Avec l'AC, la fonctionnelle universelle en fonction de λ devient

$$F^\lambda[\rho] = \min_{\Psi \rightarrow \rho} \langle \Psi | \hat{T} + \lambda \hat{V}_{ee} | \Psi \rangle = \langle \Psi_\lambda | \hat{T} + \lambda \hat{V}_{ee} | \Psi_\lambda \rangle \quad (1.82)$$

ce qui donne aussi

$$F^\lambda[\rho] = T_s[\rho] + U[\rho] + E_{xc}^\lambda[\rho] \quad (1.83)$$

où T_s est l'énergie cinétique pour des électrons non interagissant, $U[\rho]$ est l'énergie d'Hartree (l'énergie de répulsion coulombienne entre deux charges) et E_{xc}^λ est l'énergie XC qui dépend de λ . L'AC donne une définition de l'énergie de corrélation en DFT

$$E_c[\rho] = \int_0^1 d\lambda \langle \Psi_\lambda | \hat{V}_{ee} | \Psi_\lambda \rangle - \langle \Phi_{KS} | \hat{V}_{ee} | \Phi_{KS} \rangle. \quad (1.84)$$

où Φ_{KS} est un déterminant de Slater pour le système KS non interagissant, donc pour $\lambda = 0$. La définition de l'énergie de corrélation en DFT est différente de celle des méthodes post-HF (voir l'équation 1.34). On remarque que l'énergie cinétique de corrélation n'est pas présente dans l'équation précédente. Une façon pratique d'inclure T_c en DFT est d'ajouter cette énergie dans le trou d'échange-corrélation, ce qui sera décrit à la section suivante.

1.4. Trou d'échange et de corrélation

Un concept important en chimie quantique décrit les effets des interactions électroniques en DFT: le trou d'échange-corrélation (trou XC). Il va, par exemple, inclure les effets de l'exclusion de Pauli en repoussant les électrons de mêmes spins, considère la corrélation

instantanée entre les électrons et retire la répulsion d'un électron avec lui-même (erreur d'auto-interaction). Pour comprendre d'où provient le trou XC et comment l'obtenir, il faut introduire la densité de paire

$$P_2(\mathbf{x}_1, \mathbf{x}_2) = N(N-1) \int \dots \int |\Psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)|^2 d\mathbf{x}_3 \dots \mathbf{x}_N, \quad (1.85)$$

où $\mathbf{x}_i = \mathbf{r}_i \sigma_i$ et $\sigma = \alpha$ ou β . La formule précédente représente la densité de probabilité de trouver simultanément deux électrons de spins σ_1 et σ_2 à \mathbf{r}_1 et \mathbf{r}_2 . Selon la connexion adiabatique décrite à la section précédente, la densité de paire moyenne selon la force de couplage λ est

$$\bar{P}_2(\mathbf{x}_1, \mathbf{x}_2) = \int_0^1 d\lambda P_2^\lambda(\mathbf{x}_1, \mathbf{x}_2) \quad (1.86)$$

et elle montre plusieurs propriétés importantes pour la compréhension des effets quantiques en DFT. Puisque les électrons sont des fermions, deux électrons de même spin ne peuvent pas être au même endroit dans l'espace

$$\bar{P}_2(\mathbf{x}_1, \mathbf{x}_1) = 0. \quad (1.87)$$

Lorsque la corrélation électronique n'est pas considérée (donc pour $\lambda = 0$), seulement les électrons de spin opposés ne sont pas corrélés. La densité de paire, lorsque $\lambda = 0$ nous donne

$$P_2^{\lambda=0}(\mathbf{r}_1 \sigma_1, \mathbf{r}_2 \sigma_2) = \rho(\mathbf{r}_1 \sigma_1) \rho(\mathbf{r}_2 \sigma_2), \text{ où } \sigma_1 \neq \sigma_2 \quad (1.88)$$

où $\rho(\mathbf{x})$ est la densité électronique. La probabilité d'obtenir un électron à une position \mathbf{r}_1 est ainsi indépendante de la position d'un électron à \mathbf{r}_2 . Pour simplifier les équations, nous allons maintenant considérer les équations où une moyenne sur le spin est effectuée, donc la variable de spin est intégrée

$$\rho_2^\lambda(\mathbf{r}_1, \mathbf{r}_2) = \int \int d\sigma_1 d\sigma_2 P_2^\lambda(\mathbf{r}_1 \sigma_1, \mathbf{r}_2 \sigma_2). \quad (1.89)$$

Pour introduire le trou XC, on peut commencer par une intégrale sur la force de couplage λ

$$\rho_{xc}(\mathbf{r}_1, \mathbf{r}_2) = \int_0^1 d\lambda \rho_{xc}^\lambda(\mathbf{r}_1, \mathbf{r}_2), \quad (1.90)$$

avec

$$\rho_{xc}^\lambda(\mathbf{r}_1, \mathbf{r}_2) = \frac{\rho_2^\lambda(\mathbf{r}_1, \mathbf{r}_2)}{\rho(\mathbf{r}_1)} - \rho(\mathbf{r}_2). \quad (1.91)$$

Alors, le trou représente la diminution de la densité électronique à une position \mathbf{r}_2 , causé par un électron de référence à la position \mathbf{r}_1 .

Selon la connexion adiabatique, on peut montrer que le trou XC considère la contribution de la corrélation à l'énergie cinétique. Pour prouver que T_c peut être incorporé dans le trou

XC, on peut écrire

$$E_{\lambda=1} - E_{\lambda=0} = \int d\mathbf{r} \rho(\mathbf{r}) \left(v_{ext}^{\lambda=1}(\mathbf{r}) - v_{ext}^{\lambda=0}(\mathbf{r}) \right) + U[\rho] + \frac{1}{2} \int_0^1 d\lambda \int \int d\mathbf{r}_1 d\mathbf{r}_2 \frac{\rho(\mathbf{r}_1) \rho_{xc}^{\lambda}(\mathbf{r}_1, \mathbf{r}_2)}{r_{12}} \quad (1.92)$$

En utilisant $E_{\lambda=0} = T_S + \int \rho(\mathbf{r}) v_{ext}^{\lambda=0}(\mathbf{r}) d\mathbf{r}$, on obtient finalement la formule

$$E_{\lambda=1} = T_S + \int d\mathbf{r} \rho(\mathbf{r}) v_{ext}^{\lambda=1}(\mathbf{r}) + \frac{1}{2} \int \int d\mathbf{r}_1 d\mathbf{r}_2 \frac{\rho(\mathbf{r}_1) \rho(\mathbf{r}_2)}{r_{12}} + \frac{1}{2} \int \int d\mathbf{r}_1 d\mathbf{r}_2 \frac{\rho(\mathbf{r}_1) \rho_{xc}(\mathbf{r}_1, \mathbf{r}_2)}{r_{12}}, \quad (1.93)$$

ce qui correspond à l'énergie électronique d'un système où les électrons sont interagissant. Pourtant dans l'équation précédente, l'énergie cinétique est celle pour des électrons non corrélés (T_s), les effets de la corrélation à l'énergie cinétique doivent donc être inclus dans $\rho_{xc}(\mathbf{r}_1, \mathbf{r}_2)$. La connexion adiabatique est un concept qui est à la base des fonctionnelles hybrides en DFT [33] et elle sera expliquée davantage à la section 1.5.5.

C'est le chercheur Slater qui a introduit le concept de trous en chimie quantique [16, 34], en cherchant une méthode plus simple et plus rapide à calculer que HF. Cette simplification était nécessaire, considérant les grandes limitations computationnelles durant les années 1950. Le trou XC est une quantité extrêmement importante en DFT, comme le montrent les travaux de cette thèse. Outre notre groupe de recherche, on peut citer, par exemple, les travaux du groupe de Kong [35] qui a comparé le trou X exact (ce trou est connu pour un déterminant de Slater) avec des approximations courantes. Les études des propriétés des trous X, C et XC sont pertinentes afin de développer des approximations simples et efficaces pour accélérer les calculs.

L'équation décrivant l'énergie associée aux interactions entre les électrons selon la connexion adiabatique est la suivante

$$E_{ee} = \frac{1}{2} \int \int d\mathbf{r}_1 d\mathbf{r}_2 \frac{\rho(\mathbf{r}_1) \rho(\mathbf{r}_2)}{r_{12}} + \frac{1}{2} \int \int d\mathbf{r}_1 d\mathbf{r}_2 \frac{\rho(\mathbf{r}_1) \rho_{xc}(\mathbf{r}_1, \mathbf{r}_2)}{r_{12}}. \quad (1.94)$$

On remarque, à partir de l'équation précédente, que la répulsion coulombienne pour des électrons indépendants est représentée par le premier terme. C'est l'équation classique de la répulsion entre deux distributions de charge. Le deuxième terme constitue l'interaction entre la densité électronique et le trou XC et contient les phénomènes quantiques, tels que l'échange et la corrélation. L'équation 1.93 montre que la contribution de la corrélation à l'énergie cinétique est incluse dans le trou XC. Cependant, le premier terme qui représente la répulsion coulombienne classique n'est pas nul même lorsqu'il n'y a qu'un seul électron. L'électron ne doit pas interagir avec lui-même selon la DFT. Néanmoins, c'est seulement le total E_{ee} qui doit être nul, donc, pour un système à un électron, le trou XC doit être égal, mais du signe opposé à $\rho(\mathbf{r}_2)$. La théorie HF est un exemple qui respecte cette condition et elle ne souffre alors pas de l'erreur d'auto-interaction unielectronique (*self-interaction error* SIE). Le modèle construit en 1951 par Slater [34] souffre de la SIE, tel que la majorité

Tab. I-1. Résumé de plusieurs propriétés mathématiques des trous d'échange-corrélation, d'échange et de corrélation.

Trous	Propriétés mathématiques
$\rho_{xc}(\mathbf{r}_1, \mathbf{r}_2)$	$-1 = \int d\mathbf{r}_2 \rho_{xc}(\mathbf{r}_1, \mathbf{r}_2)$ $\rho_{xc}(\mathbf{r}_1, \mathbf{r}_2) = \rho_c(\mathbf{r}_1, \mathbf{r}_2) + \rho_x(\mathbf{r}_1, \mathbf{r}_2)$ $\rho_{xc}(\mathbf{r}_1, \mathbf{r}_2) = f_c(\mathbf{r}_1, \mathbf{r}_2) \rho_x(\mathbf{r}_1, \mathbf{r}_2)$
$\rho_x(\mathbf{r}_1, \mathbf{r}_2)$	$\rho_{x,\sigma}(\mathbf{r}_1, \mathbf{r}_2) = -\frac{ \sum_i \phi_{i,\sigma}^*(\mathbf{r}_1) \phi_{i,\sigma}(\mathbf{r}_2) ^2}{\rho_\sigma(\mathbf{r}_1)}$ $\rho_{x,\sigma}(\mathbf{r}_1, \mathbf{r}_2) \leq 0 \forall \mathbf{r}_1, \mathbf{r}_2$ $-1 = \int d\mathbf{r}_2 \rho_{x,\sigma}(\mathbf{r}_1, \mathbf{r}_2)$ $\rho_{x,\sigma}(\mathbf{r}_1, \mathbf{r}_1) = -\rho_\sigma(\mathbf{r}_1)$ $\rho(\mathbf{r}_1) \rho_x(\mathbf{r}_1, \mathbf{r}_2) = \rho_\alpha(\mathbf{r}_1) \rho_{x,\alpha}(\mathbf{r}_1, \mathbf{r}_2) + \rho_\beta(\mathbf{r}_1) \rho_{x,\beta}(\mathbf{r}_1, \mathbf{r}_2)$ $\rho_{x,\sigma}(\mathbf{r}, u) = -\rho_\sigma(\mathbf{r}) - Q_\sigma(\mathbf{r}) u^2 - T_\sigma(\mathbf{r}) u^4 + \dots$
$\rho_c(\mathbf{r}_1, \mathbf{r}_2)$	$0 = \int d\mathbf{r}_2 \rho_c(\mathbf{r}_1, \mathbf{r}_2)$ $\rho_c(\mathbf{r}_1, \mathbf{r}_2) \in \mathbb{R}$ $\rho(\mathbf{r}_1) \rho_c(\mathbf{r}_1, \mathbf{r}_2) = \rho_\alpha(\mathbf{r}_1) (\rho_{c,\alpha}(\mathbf{r}_1, \mathbf{r}_2) + \rho_{c,\alpha\beta}(\mathbf{r}_1, \mathbf{r}_2)) + \rho_\beta(\mathbf{r}_1) (\rho_{c,\beta}(\mathbf{r}_1, \mathbf{r}_2) + \rho_{c,\beta\alpha}(\mathbf{r}_1, \mathbf{r}_2))$
Tous	$\rho_{xc/x/c}(\mathbf{r}, u) = \int \frac{d\Omega_u}{4\pi} \rho_{xc/x/c}(\mathbf{r}, \mathbf{r} + \mathbf{u})$

des fonctionnelles courantes en DFT. L'élimination et la quantification de la SIE sont des sujets importants en DFT [36, 37], puisqu'elle peut, par exemple, diminuer l'exactitude des prédictions des barrières d'activation [38].

1.4.1. Propriétés du trou d'échange-corrélation.

Le trou XC n'est pas connu exactement par une formule mathématique simple. Toutefois, le trou XC a plusieurs particularités et propriétés qui permettent de mieux le comprendre, ce qui facilite la construction des approximations au trou. Le tableau I-1 résume la plupart des propriétés des trous XC, X et C dont nous allons discuter. Par exemple, le trou XC peut être séparé en la somme d'un trou X et un trou C

$$\rho_{xc}(\mathbf{r}_1, \mathbf{r}_2) = \rho_x(\mathbf{r}_1, \mathbf{r}_2) + \rho_c(\mathbf{r}_1, \mathbf{r}_2). \quad (1.95)$$

Si l'on considère le spin, l'interaction d'échange est présente seulement pour les électrons de mêmes spins

$$\rho(\mathbf{r}_1) \rho_x(\mathbf{r}_1, \mathbf{r}_2) = \rho_\alpha(\mathbf{r}_1) \rho_{x,\alpha}(\mathbf{r}_1, \mathbf{r}_2) + \rho_\beta(\mathbf{r}_1) \rho_{x,\beta}(\mathbf{r}_1, \mathbf{r}_2), \quad (1.96)$$

tandis que la corrélation est aussi présente pour les électrons de spins opposés

$$\rho(\mathbf{r}_1)\rho_c(\mathbf{r}_1,\mathbf{r}_2) = \rho_\alpha(\mathbf{r}_1) (\rho_{c,\alpha}(\mathbf{r}_1,\mathbf{r}_2) + \rho_{c,\alpha\beta}(\mathbf{r}_1,\mathbf{r}_2)) + \rho_\beta(\mathbf{r}_1) (\rho_{c,\beta}(\mathbf{r}_1,\mathbf{r}_2) + \rho_{c,\beta\alpha}(\mathbf{r}_1,\mathbf{r}_2)). \quad (1.97)$$

De plus, comme l'opérateur pour calculer l'énergie de répulsion électronique est de la forme $1/r_{12}$, il ne dépend pas de l'orientation entre les vecteurs \mathbf{r}_1 et \mathbf{r}_2 , mais seulement de la magnitude. Il est possible d'obtenir un trou XC où une moyenne sphérique a été effectuée. Pour ce faire, il est plus simple d'introduire un nouveau vecteur \mathbf{u} et l'on obtient

$$\rho_{xc}(\mathbf{r},u) = \int \frac{d\Omega_u}{4\pi} \rho_{xc}(\mathbf{r},\mathbf{r} + \mathbf{u}) \quad (1.98)$$

où Ω_u représente toutes les orientations du vecteur \mathbf{u} et u est la distance à partir du point \mathbf{r} . La moyenne sphérique indique que si l'on se place à la position de référence \mathbf{r} , les valeurs du trou XC à chaque distance u seront des moyennes de ce qui est observé à chaque orientation autour de la référence. La moyenne sphérique simplifie l'équation du trou XC en éliminant deux variables angulaires.

La somme entre le trou X et C n'est pas la définition unique pour obtenir le trou XC, c'est seulement le trou total qui est important pour trouver l'énergie XC en pratique. Une alternative à l'équation 1.95 est un facteur de corrélation multiplicatif, une approche développée par le groupe Ernzerhof [39, 40]

$$\rho_{xc}(\mathbf{r}_1,\mathbf{r}_2) = f_c(\mathbf{r}_1,\mathbf{r}_2)\rho_x(\mathbf{r}_1,\mathbf{r}_2) \quad (1.99)$$

où $f_c(\mathbf{r}_1,\mathbf{r}_2)$ se nomme le facteur de corrélation. Lors de la construction d'approximations au trou XC, un facteur multiplicatif comme le facteur de corrélation est plus simple que de construire un trou C, car le trou X est une fonction non locale. Donc, pour avoir un trou C compatible avec le trou X, ainsi qui mène à de bonnes prédictions, il faut aussi qu'il soit non local, ce qui est difficile à construire en pratique due à la contrainte de le rendre compatible avec le trou X lors de l'addition. Concernant le facteur de corrélation, il est plus facile d'éliminer la composante non locale du trou X en utilisant une fonction décroissante comme une fonction gaussienne.

La construction de trou C approximatif est extrêmement difficile, il y a très peu d'articles sur le sujet. Nous pouvons par exemple citer ceux du groupe de Perdew concernant le trou C du gaz d'électrons homogène [41, 42]. Pour le facteur de corrélation, notre groupe a déjà publié plusieurs articles avec des fonctionnelles qui ont des résultats équivalents ou supérieurs aux méthodes courantes en DFT [40, 43–46].

Le trou X est particulièrement intéressant à étudier, car l'équation est relativement simple lorsque la fonction d'onde est un unique déterminant de Slater. Le trou X exact pour un spin σ est donné par l'équation

$$\rho_{x,\sigma}(\mathbf{r}_1,\mathbf{r}_2) = -\frac{|\sum_i \phi_{i,\sigma}^*(\mathbf{r}_1)\phi_{i,\sigma}(\mathbf{r}_2)|^2}{\rho_\sigma(\mathbf{r}_1)}. \quad (1.100)$$

L'équation précédente a permis à Slater de construire la première approximation au trou X. Il est essentiel d'avoir une approximation précise de l'énergie X, puisqu'elle compose la majorité de l'énergie XC en éliminant l'erreur d'auto-interaction [16]. Une propriété mathématique du trou X est sa normalisation à tous les points \mathbf{r}_1

$$-1 = \int d\mathbf{r}_2 \rho_{x,\sigma}(\mathbf{r}_1, \mathbf{r}_2) \quad (1.101)$$

ce qui correspond au retrait d'un électron. Cette propriété est prouvée à partir de l'intégrale de la densité de paire qui doit être égale à $N(N-1)$. La normalisation du trou X est une contrainte universelle qui est fortement utile pour construire des approximations à ce dernier. De plus, afin que la densité de paire, pour des spins identiques, soit nulle à l'origine (équation 1.87), donc pour respecter le principe de Pauli, le trou d'échange doit avoir la valeur de la densité à la position d'origine

$$\rho_{x,\sigma}(\mathbf{r}_1, \mathbf{r}_1) = -\rho_\sigma(\mathbf{r}_1). \quad (1.102)$$

D'ailleurs, le développement en série de Taylor de la moyenne sphérique du trou X est connu jusqu'au terme de quatrième ordre

$$\rho_{x,\sigma}(\mathbf{r}, u) = -\rho_\sigma(\mathbf{r}) - Q_\sigma(\mathbf{r})u^2 - T_\sigma(\mathbf{r})u^4 + \dots, \quad (1.103)$$

où

$$Q_\sigma(\mathbf{r}) = \frac{1}{6} \left[\nabla^2 \rho_\sigma(\mathbf{r}) - 2\tau_\sigma(\mathbf{r}) + \frac{1}{2} \frac{(\nabla \rho_\sigma(\mathbf{r}))^2}{\rho_\sigma(\mathbf{r})} \right] \quad (1.104)$$

avec

$$\tau_\sigma(\mathbf{r}) = \sum_i^{occ} \frac{1}{2} |\nabla \psi_{i,\sigma}(\mathbf{r})|^2 \quad (1.105)$$

et

$$T_\sigma(\mathbf{r}) = \frac{1}{5! \rho_\sigma(\mathbf{r})} \left[2\rho_\sigma(\mathbf{r}) \sum_j^{occ} \psi_{j,\sigma}(\mathbf{r}) \nabla^4 \psi_{j,\sigma}(\mathbf{r}) + 4\nabla \rho_\sigma(\mathbf{r}) \cdot \sum_j^{occ} \psi_{j,\sigma}(\mathbf{r}) \nabla^3 \psi_{j,\sigma}(\mathbf{r}) + \frac{1}{2} (\nabla^2 \rho_\sigma(\mathbf{r}) - 2\tau_\sigma(\mathbf{r}))^2 + 4 \sum_i^{occ} \psi_{i,\sigma}(\mathbf{r}) \psi_{i,\sigma}^{v6}(\mathbf{r}) \cdot \sum_j^{occ} \psi_{j,\sigma}(\mathbf{r}) \psi_{j,\sigma}^{v6}(\mathbf{r}) \right] \quad (1.106)$$

avec

$$\psi^{v6}(\mathbf{r}) = (\psi_{xx}(\mathbf{r}), \psi_{yy}(\mathbf{r}), \psi_{zz}(\mathbf{r}), \sqrt{2}\psi_{xy}(\mathbf{r}), \sqrt{2}\psi_{xz}(\mathbf{r}), \sqrt{2}\psi_{yz}(\mathbf{r})), \quad (1.107)$$

où les indices représentent une dérivée par rapport aux coordonnées. Le terme $Q_\sigma(\mathbf{r})$ a été dérivé mathématiquement par Becke [47, 48] et le terme de quatrième ordre $T_\sigma(\mathbf{r})$ a été obtenu par notre groupe en 2017 [49]. Les termes impairs sont nuls dans le développement en série de Taylor, dû à la moyenne sphérique. $Q_\sigma(\mathbf{r})$ est utilisé, entre autres, pour obtenir le trou X Becke-Roussel [48], ce qui sera expliqué à la section 1.5.4. $T_\sigma(\mathbf{r})$ est une propriété beaucoup moins étudiée que $Q_\sigma(\mathbf{r})$, au meilleur de nos connaissances, seulement notre étude de 2017 [49] s'est intéressée à $T_\sigma(\mathbf{r})$. Selon l'article, $T_\sigma(\mathbf{r})$ est utile pour améliorer la description du trou X des molécules, ainsi, il permet d'améliorer le calcul de l'énergie d'échange des

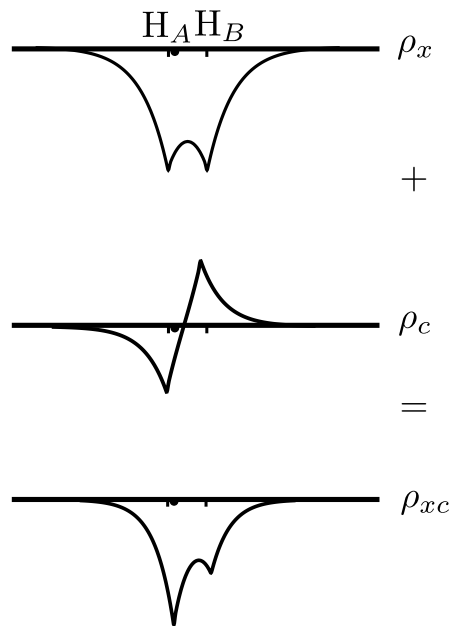


Fig. 1.4. Esquisses des trous d'échange (X), de corrélation (C) et total pour la molécule de H_2 à sa longueur de lien d'équilibre. La position de l'électron de référence est représentée par le point qui est près de la position de l'atome H_A .

molécules. Cependant, le développement de modèle qui utilise $T_\sigma(\mathbf{r})$ est difficile. Le projet de l'article 3 va s'intéresser davantage à $T_\sigma(\mathbf{r})$, en combinant cette variable avec l'apprentissage automatique, pour créer une fonctionnelle qui s'approche de la qualité des prédictions d'une fonctionnelle hybride, sans même avoir besoin de calculer E_x^{exact} .

Finalement, la dernière propriété du trou X que nous allons discuter est que le trou X est strictement négatif, tel que l'on peut observer dans l'équation 1.100.

Concernant le trou C, il n'est pas connu exactement sous une forme analytique simple, donc il est plus difficile à étudier en pratique. La valeur de son intégrale dans l'espace doit être égale à 0, car le trou X intègre à -1 tout comme le trou XC

$$0 = \int d\mathbf{r}_2 \rho_c(\mathbf{r}_1, \mathbf{r}_2). \quad (1.108)$$

Le trou XC est particulièrement utile pour étudier la corrélation statique, car ce type de corrélation est beaucoup moins intuitive à comprendre que la corrélation dynamique. Il est difficile de visualiser comment des énergies de niveaux presque dégénérés peuvent causer de la corrélation. Lorsqu'il y a de la corrélation statique importante dans un système, les trous X et C ont une partie à longue portée, donc non locale [50, 51]. Par exemple, le trou XC de la molécule de H_2 dissociée est assez local (voir la figure 1.5 et la référence [51]), car le trou XC est localisé autour de l'électron de référence. Cependant, la moitié du trou X est présent sur l'autre atome. Le trou C doit donc lui aussi avoir une partie à longue portée pour

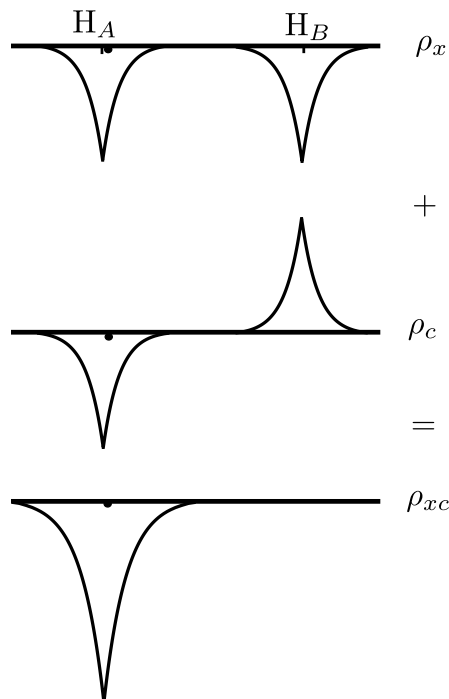


Fig. 1.5. Esquisses des trous d'échange (X), de corrélation (C) et total pour la molécule de H_2 étirée. La position de l'électron de référence est près de l'atome H_A .

annuler celle de l'échange sur l'atome dissocié pour que le trou total XC soit local. Ainsi, la portée du trou X et C détecte la présence de corrélation statique. Cette observation est à la base de la fonctionnelle B05 de Becke [52] qui ajoute de la corrélation statique en fonction de la normalisation locale d'un trou X approximatif. Lorsque cette normalisation locale est de -0.5 plutôt que la valeur attendue de -1, la corrélation statique est alors significative. Comme mentionné dans la section 1.1.4, la théorie MCSCF calcule l'énergie de corrélation statique, mais l'approche de Becke est beaucoup plus rapide. Nous allons discuter davantage de ces deux approches pour calculer la corrélation statique dans l'article 1.

En résumé, le trou XC est une propriété utile en WFT et DFT pour mieux comprendre les propriétés quantiques de la matière. Cette meilleure compréhension est non seulement pertinente d'un point de vue fondamental, mais elle permet de construire des approximations qui accélèrent les calculs comme Slater l'a fait en 1951 [34].

1.5. Approximations de l'énergie d'échange-corrélation

Selon l'approche KS, le seul terme inconnu à approximer est l'énergie d'échange-corrélation (XC). Il existe deux philosophies principales concernant la construction de fonctionnelles à l'énergie XC en DFT [16, 53, 54]. La première, celle qui va nous intéresser majoritairement dans cette thèse, est le développement d'approximations qui satisfont des contraintes physiques et mathématiques. Quelques paramètres ajustables peuvent

être ajoutés, afin de les optimiser pour reproduire des valeurs provenant de bases de données théoriques ou expérimentales. La valeur obtenue par ces paramètres ajustables doit cependant être possible à analyser d'un point de vue physique et chimique. Cette approche va généralement mener à des approximations qui ont des erreurs systématiques et qui peuvent être rationalisées [53].

La deuxième philosophie en développement de fonctionnelles est considérée comme étant empirique. Une fonction flexible avec plusieurs paramètres ajustables est optimisée afin de reproduire une grande quantité de valeurs provenant des bases de données expérimentales ou générées par la WFT. La signification physique des valeurs des paramètres n'est pas importante, les modèles doivent reproduire les données d'entraînement avec une bonne exactitude. Grâce à cette optimisation, cette approche va permettre d'obtenir des résultats supérieurs aux méthodes non empiriques pour des molécules qui ressemblent à celles qui ont été utilisées lors de l'entraînement. Cependant, les erreurs sont moins systématiques que pour les méthodes non empiriques, donc il est plus difficile de rationaliser les erreurs obtenues. Les fonctionnelles dites empiriques peuvent aussi être inférieures à celles non empiriques pour des molécules qui sont trop différentes des molécules considérées lors de l'entraînement. De plus, les fonctionnelles sont généralement optimisées pour reproduire des bases de données contenant de l'information reliée à l'énergie des molécules, donc d'autres propriétés comme la densité électronique sont souvent ignorées lors de l'optimisation des paramètres [55]. Une fonctionnelle empirique peut fournir d'excellentes prédictions pour certaines propriétés, au coût d'avoir des résultats inférieurs pour d'autres propriétés. Un article publié en 2017 dans le journal *Science* par Medvedev et co-auteurs [55] a montré que les fonctionnelles empiriques récentes donnent de bons résultats pour reproduire des bases de données en thermochimie, mais le calcul de la densité électronique donne des résultats détériorés. Selon Google Scholar, cet article a plus de 800 citations en 5 ans, démontrant l'impact de cet article dans le domaine du développement de fonctionnelles en DFT. Nous avons effectué une analyse de la densité électronique obtenue par une de nos fonctionnelles dans l'article 3. Notre fonctionnelle utilise un modèle flexible comme les réseaux de neurones, il est donc essentiel de s'assurer que les prédictions obtenues sont raisonnables pour différentes propriétés des molécules.

Ainsi, comparativement à la WFT, la grande difficulté lors du développement d'approximations en DFT est qu'il n'y a pas de façon systématique pour améliorer les fonctionnelles. En WFT, la méthode de configuration d'interaction complète donne une énergie exacte de l'équation de Schrödinger indépendante du temps, mais les calculs sont trop lents en pratique. En DFT, la fonctionnelle exacte est inconnue, ce qui complique le développement de nouvelles approximations.

Une tentative de classification des fonctionnelles en DFT se nomme "l'échelle de Jacob" [56] et elle est montrée à la figure 1.6. Plus on monte dans l'échelle, plus l'exactitude de la méthode augmente en général (la qualité des résultats peut dépendre des molécules

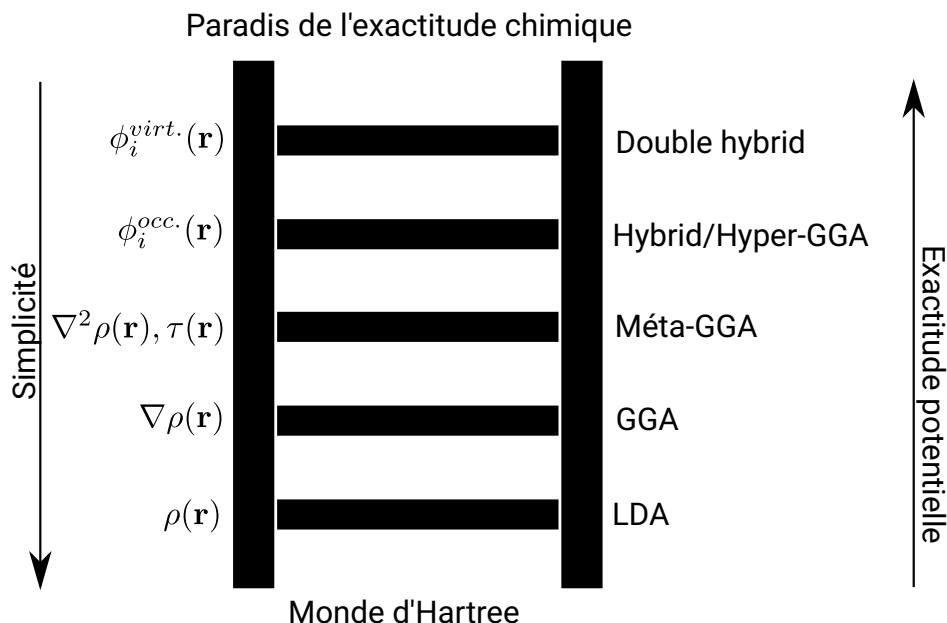


Fig. 1.6. Échelle de Jacob de la DFT [56]. Les approximations des niveaux supérieurs vont généralement inclure les ingrédients des niveaux inférieurs. LDA est l'approximation de la densité locale et GGA est l'approximation du gradient généralisé.

étudiées), mais la complexité augmente aussi. Le premier niveau contient la toute première approximation qui a été développée en KS-DFT [20] et elle est basée sur le modèle du gaz d'électron homogène (HEG). Le seul ingrédient qu'elle utilise est la densité électronique $\rho(\mathbf{r})$ à chaque position. Cette approximation est généralement nommée l'approximation de la densité locale (*local density approximation* LDA). Le deuxième niveau d'approximation ajoute le gradient de la densité $\nabla\rho(\mathbf{r})$ comme ingrédient et se nomme *Generalized Gradient Approximation* (GGA). L'approximation GGA la plus populaire est celle de Perdew-Burke-Ernzerhof (PBE) [57]. Ensuite, les Méta-GGA ajoutent la densité d'énergie cinétique $\tau(\mathbf{r})$ et parfois le laplacien de la densité $\nabla^2\rho(\mathbf{r})$. Une Méta-GGA récente qui gagne en popularité est SCAN [58]. Ensuite, les fonctionnelles hybrides et les "Hyper-GGA" vont inclure de l'information provenant des orbitales occupées. Les orbitales permettent d'obtenir l'énergie d'échange exacte, qui est calculée de façon analogue à celle de HF. La méthode hybride la plus connue est B3LYP [59, 60], mais dans cette thèse nous allons aussi nous intéresser à la fonctionnelle hybride de PBE: PBEh [61, 62]. Finalement, la dernière marche de l'échelle inclut les orbitales virtuelles $\phi_i^{virt.}$. Il y a par exemple les doubles hybrides comme B2BLYP [63] dont une partie de l'énergie de corrélation provient de la théorie de perturbation. Un exemple de ce type de fonctionnelles qui a été développé par notre groupe de recherche est CFXStatic [64], qui utilise l'énergie de corrélation statique provenant de MCSCF.

Nous allons maintenant discuter de l'inclusion du spin dans les équations des fonctionnelles, puis détailler les types de fonctionnelles les plus importants à comprendre pour cette thèse.

1.5.1. *Les fonctionnelles et le spin.*

Selon l'approche KS, seulement la densité électronique totale est nécessaire pour calculer l'énergie électronique exactement. En l'absence de champ magnétique, l'utilisation des spin-densités n'est pas essentielle. Cependant, comme l'énergie XC doit être approximée, la considération du spin des électrons permet plus de flexibilité aux modèles en ajoutant davantage de paramètres à optimiser. Il est particulièrement important d'inclure le spin aux modèles pour bien décrire les molécules ayant des électrons non appariés [1].

Il est possible de calculer la densité électronique totale à partir des spins densités

$$\rho(\mathbf{r}) = \rho_\alpha(\mathbf{r}) + \rho_\beta(\mathbf{r}) \quad (1.109)$$

et une propriété fréquemment utilisée en DFT est la polarisation de spin

$$\zeta(\mathbf{r}) = \frac{\rho_\alpha(\mathbf{r}) - \rho_\beta(\mathbf{r})}{\rho(\mathbf{r})}, \quad (1.110)$$

où $\zeta(\mathbf{r})$ est nul à toutes les positions pour une molécule non polarisée, tandis que la valeur est de 1 pour un système unielectronique comme l'hydrogène.

Finalement, dans la littérature, les équations en KS-DFT sont représentées sous la forme spin-non polarisée ou spin-polarisée [16]. Puisque la séparation entre les spins est beaucoup plus simple pour l'échange (l'échange existe seulement pour des électrons de mêmes spins) c'est cette énergie qui est décrite sous les deux formes différentes. La relation entre les deux formalismes est [65]

$$E_x[\rho_\alpha, \rho_\beta] = \frac{1}{2}(E_x[2\rho_\alpha] + E_x[2\rho_\beta]) \quad (1.111)$$

où la partie de gauche est l'énergie d'échange lorsque les équations sont sous forme spin-polarisée et à droite elles sont sous la forme non polarisée. Dans ce texte, nous allons majoritairement écrire les équations selon le formalisme non spin-polarisé. Lorsque ce ne sera pas le cas, le spin sera écrit explicitement.

1.5.2. *Approximation de la densité locale.*

Le modèle du gaz d'électrons homogène (HEG) consiste en un nuage électronique et une charge positive uniformément distribuée dans l'espace. Ce modèle est à la base de l'approximation de la densité locale (LDA) et de l'approximation de la spin-densité locale (LSDA) lorsque la composante de spin est considérée. La LDA est la première approximation de la KS-DFT et a été construite par Kohn et Sham en 1965 pour calculer l'énergie XC [20].

L'énergie de la LDA ne dépend que de la densité électronique à chaque position dans l'espace

$$E_{xc}^{LDA} = \int d\mathbf{r} \rho(\mathbf{r}) \epsilon_{xc}^{LDA}(\rho(\mathbf{r})). \quad (1.112)$$

Ainsi, l'énergie XC d'une molécule pour une position \mathbf{r} est approximée comme étant équivalente à celle d'un HEG ayant le même $\rho(\mathbf{r})$. Comme pour la majorité des fonctionnelles, l'énergie XC est séparée en une contribution pour l'échange et la corrélation

$$E_{xc}^{LDA} = E_x^{LDA} + E_c^{LDA}. \quad (1.113)$$

L'énergie d'échange est donnée par une formule très simple

$$E_x^{LDA}[\rho] = - \int d\mathbf{r} \rho(\mathbf{r}) \frac{3}{4} \left(\frac{3\rho(\mathbf{r})}{\pi} \right)^{1/3}, \quad (1.114)$$

c'est l'équation qui a été développée par Dirac et dont nous avons discuté lors de la description du modèle Thomas-Fermi-Dirac à la section 1.2. L'équation pour décrire l'énergie de corrélation du HEG est plus compliquée, il n'y a pas une formule analytique simple qui a été dérivée, donc les théoriciens ont développé plusieurs façons de la calculer. Les formules les plus populaires utilisent une équation mathématique pour reproduire les résultats de haute qualité de Cerperley et Alder [66] provenant de la théorie de Monte-Carlo quantique, ainsi que les limites connues du comportement de la corrélation du HEG à forte et faible densité électronique. Parmi les équations qui ont été développées, les plus populaires proviennent de Perdew et Wang [67] et Volsko, Wilk et Nusair (VWN) [68]. Les formules mathématiques sont longues et ne seront pas retranscrites ici.

La LDA a eu du succès en physique du solide, mais son exactitude est généralement insuffisante pour être appliquée aux atomes et aux molécules, qui sont très loin d'être des systèmes avec une densité uniforme dans l'espace. La LDA a la tendance à fortement surestimer l'énergie d'atomisation [2, 16]. La MAE calculée pour les énergies d'atomisation de la base de données G2/97 [15] est de 83 kcal/mol, ce qui est environ la moitié de HF. Toutefois, c'est tout de même un résultat largement inférieur à une fonctionnelle hybride comme PBEh dont la MAE est de 4 kcal/mol pour la même base de données. La LDA est meilleure pour prédire l'énergie des molécules que celles des atomes, puisque la densité d'une molécule est un peu plus près d'un système homogène que celle d'un atome [69]. Ainsi, la LDA surestime la force des liens chimique en prédisant une énergie plus élevée pour les atomes et l'énergie d'atomisation est trop grande. La LDA est une fonctionnelle sans aucun paramètre empirique, donc il est possible d'analyser ses erreurs qui ont un comportement systématique.

De plus, le trou d'échange du HEG est connu exactement, le trou X réduit ($\rho_x(\mathbf{r}, u) = \rho(\mathbf{r}) J_x(\mathbf{r}, y)$) est

$$J_x^{HEG}(y) = - \left(\frac{3}{\sqrt{2}y^3} (\sin y - y \cos y) \right)^2 \quad (1.115)$$

où $y = k_F u$, $k_F = (3\pi^2\rho)^{1/3}$ et u est la distance autour de l'électron de référence. L'utilisation de $k_F u$ rend le trou indépendant de $\rho(\mathbf{r})$, ce qui facilite l'analyse graphique. La connaissance d'un trou X analytique est très utile pour comprendre ses propriétés afin de construire des approximations plus complexes [65, 69].

1.5.3. Approximation du gradient généralisé.

La LDA utilise seulement $\rho(\mathbf{r})$ comme ingrédient, donc la suite logique est de considérer le gradient de la densité $\nabla\rho(\mathbf{r})$ pour construire des approximations à l'énergie XC. Contrairement à la LDA, il y a différentes familles de modèles qui utilisent $\nabla\rho(\mathbf{r})$. L'approche la plus populaire se nomme l'approximation du gradient généralisé (GGA) et plusieurs groupes de recherche ont construits des fonctionnelles XC de GGA. L'approximation de l'expansion du gradient (GEA) a eu peu de succès pour la prédiction de propriétés moléculaires et ne sera pas discutée dans ce texte, mais le développement des GEA a tout de même été utile pour la construction des GGA [1].

Les GGA ont une exactitude suffisante pour être utilisée en chimie, ce qui a permis à la DFT de devenir la méthode de choix en chimie computationnelle [16, 54], surtout grâce aux fonctionnelles hybrides qui seront décrites à la section 1.5.5. La formule générale pour les GGA est la suivante

$$\epsilon_{xc}^{GGA}(\mathbf{r}) = F_{xc}(\rho(\mathbf{r}), \zeta(\mathbf{r}), \nabla\rho(\mathbf{r})) \epsilon_{xc}^{LDA}(\rho(\mathbf{r}), \zeta(\mathbf{r})) \quad (1.116)$$

où $F_{xc}(\rho(\mathbf{r}), \zeta(\mathbf{r}), \nabla\rho(\mathbf{r}))$ se nomme le facteur de renforcement XC. Ce facteur utilise le gradient de la densité pour moduler l'énergie XC provenant du gaz d'électrons homogène. Lorsque le gradient de la densité est nul, une contrainte qui est utile est $F_{xc}(\rho(\mathbf{r}), \zeta(\mathbf{r}), \nabla\rho(\mathbf{r}) = 0) = 1$, afin de correctement reproduire la limite du HEG.

En pratique, comme pour la LDA, l'énergie XC est séparée en des termes d'échange et de corrélation. L'énergie X est donnée par la formule

$$E_x^{GGA} = \int \rho(\mathbf{r}) F_x(s(\mathbf{r})) \epsilon_x^{HEG}(\mathbf{r}) \quad (1.117)$$

où s se nomme le gradient réduit $s(\mathbf{r}) = \frac{\nabla\rho(\mathbf{r})}{\rho(\mathbf{r})^{4/3}}$ et $F_x(s(\mathbf{r}))$ est le facteur de renforcement pour l'échange. L'avantage d'utiliser $s(\mathbf{r})$ plutôt que le gradient est que $s(\mathbf{r})$ est sans unité. La valeur de $s(\mathbf{r})$ est grande lorsque le gradient est grand et/ou la densité est faible, ce qui correspond aux régions de faible densité loin de l'atome ou près du cœur de l'atome [70]. À l'opposé, la valeur de $s(\mathbf{r})$ est petite lorsque le gradient est faible comme dans la région d'un lien covalent ou lorsque la densité est grande comme dans le noyau [70].

Comme mentionné précédemment, une limite importante que les fonctionnelles peuvent satisfaire est que $F_{xc}(\rho(\mathbf{r}), \zeta(\mathbf{r}), \nabla\rho(\mathbf{r}) = 0) = 1$. Cette limite reproduit le HEG lorsque le gradient est nul. Un exemple d'une fonctionnelle GGA pour l'échange est celui de PBE [57]

dont l'équation est donnée par

$$F_x^{PBE}(s(\mathbf{r})) = 1 + \kappa - \frac{\kappa}{1 + \mu s(\mathbf{r})^2/\kappa} \quad (1.118)$$

où $\kappa = 0.804$ et $\mu = 0.21951$. Les paramètres sont trouvés en satisfaisant des contraintes physiques, comme la réponse linéaire du HEG ($F_x(s(\mathbf{r})) \rightarrow 1 + \mu s(\mathbf{r})^2$ lorsque $s(\mathbf{r}) \rightarrow 0$) et la limite Lieb-Oxford [71]

$$E_x \geq -1.679 \int d\mathbf{r} \rho^{4/3}(\mathbf{r}). \quad (1.119)$$

PBE est une méthode non empirique, tandis qu'un exemple de fonctionnelle qui contient un paramètre empirique est celle de Becke de 1988 [72, 73] (B88)

$$\epsilon_x^{B88}(\mathbf{r}) = \epsilon_x^{HEG}(\mathbf{r}) + \Delta\epsilon_x^{B88}(\mathbf{r}) \quad (1.120)$$

$$\Delta\epsilon_x^{B88}(\mathbf{r}) = -\beta^{B88} \rho^{1/3}(\mathbf{r}) \frac{s(\mathbf{r})^2}{1 + 6\beta^{B88}s(\mathbf{r}) \sinh^{-1}(2^{1/3}s(\mathbf{r}))} \quad (1.121)$$

où le paramètre β^{B88} est optimisé en minimisant l'erreur sur le calcul des énergies d'échange des gaz nobles. Tout comme pour la LDA, la fonctionnelle de corrélation GGA est beaucoup plus compliquée que celle de l'échange. Dans le cas de PBE, à l'instar de l'échange, la formule et ses paramètres sont trouvés en satisfaisant diverses contraintes et limites physiques. Cependant, une limitation importante de l'énergie de corrélation calculée par PBE est qu'elle n'est pas nulle pour l'hydrogène, cette fonctionnelle souffre donc d'une erreur d'autocorrélation. Le type d'approximation de la section suivante va permettre d'éliminer cette erreur.

Les MAE pour les énergies d'atomisation des GGA ont montré des résultats assez intéressants pour attirer l'attention des chimistes. Pour la base de données d'énergies d'atomisation G2/97 [15], la fonctionnelle PBE a une MAE de 17 kcal/mol [16]. Il faut mentionner que PBE est surtout utilisée pour les solides. RevPBE [74] est une version révisée de PBE optimisée pour les molécules qui donne une MAE de 8 kcal/mol pour le calcul des énergies d'atomisation de la base de données W4-11 [75]. Comme pour la LDA, les GGA non empiriques ont tendance à surestimer l'énergie d'atomisation [69].

Concernant le trou XC des GGA, il existe un trou X modèle pour reproduire l'énergie X de PBE qui a été construit par Ernzerhof et Perdew en 1998 [65]. Tout comme pour PBE, il est non empirique et reproduit des limites physiques et mathématiques, comme le respect de la limite du HEG lorsque le gradient est nul. À propos du trou C des GGA, il n'y a que le facteur de corrélation développé par notre groupe [39, 40] qui l'approxime au meilleur de nos connaissances.

1.5.4. Méta-approximation du gradient généralisé.

Pour donner suite aux GGA, il serait intuitif d'ajouter le laplacien de la densité $\nabla^2\rho$ comme ingrédient aux fonctionnelles. Néanmoins, en pratique, la très grande majorité des

Méta-GGA utilisent plutôt la densité d'énergie cinétique

$$\tau(\mathbf{r}) = \sum_i^{occ.} \frac{1}{2} |\nabla \psi_i(\mathbf{r})|^2, \quad (1.122)$$

car $\tau(\mathbf{r})$ est généralement numériquement plus stable à calculer que le laplacien de la densité. Le laplacien de la densité diverge au noyau et il est possible d'approximer $\nabla^2 \rho(\mathbf{r})$ à l'aide de $\tau(\mathbf{r})$ [76]. L'une des utilités de ce terme provient du fait qu'il détecte les régions à un électron, avec

$$\alpha^{MGGA}(\mathbf{r}) = \frac{\tau(\mathbf{r}) - \tau^W(\mathbf{r})}{\tau^{unif.}(\mathbf{r})} \quad (1.123)$$

où $\tau^W(\mathbf{r}) = |\nabla \rho(\mathbf{r})|^2 / (8\rho(\mathbf{r}))$ et $\tau^{unif.}(\mathbf{r}) = (3/10)(3\pi^2)^{2/3} \rho(\mathbf{r})^{5/3}$. Si $\alpha^{MGGA}(\mathbf{r}) = 1$, la région est équivalente à un gaz d'électron homogène. Si $\alpha^{MGGA}(\mathbf{r}) = 0$, la région est comme un système unielectronique ou l'équivalent de deux électrons de spins opposés.

Les fonctionnelles LDA et PBE ont une énergie de corrélation non nulle pour l'hydrogène. Ainsi, les Méta-GGA avec l'ingrédient $\tau(\mathbf{r})$ corrigent ce problème. Du côté des fonctionnelles non empiriques, la méta-GGA de Tao-Perdew-Staroverov-Scuseria (TPSS) [77] est populaire et une autre qui gagne en popularité depuis 2015 se nomme SCAN [58]. SCAN satisfait 17 contraintes physiques et mathématiques connues en DFT et montre d'excellents résultats en chimie et en physique du solide. Cependant, en montant les niveaux de l'échelle de Jacob, les équations qui définissent les fonctionnelles deviennent de plus en plus compliquées. Elles ne seront donc pas retranscrites ici.

Une approximation Méta-GGA non empirique, uniquement pour l'énergie d'échange, est le trou X Becke-Roussel (BR) [48]. Le trou X BR est très important à comprendre pour cette thèse. Pour dériver l'équation, il faut commencer avec une orbitale hydrogénoïde 1s d'un spin σ ($\rho_{x,\sigma}(\mathbf{r},u) = \sum_{\sigma} \frac{\rho_{\sigma}(\mathbf{r})}{\rho} \rho_{x,\sigma}(\mathbf{r},u)$) et effectuer une moyenne sphérique

$$\rho_{x,\sigma}^H(\mathbf{r},u) = \frac{1}{4\pi} \int d\Omega_u \rho_{H,\sigma}(\mathbf{r} + \mathbf{u}) \quad (1.124)$$

où $d\Omega_u$ signifie une intégration dans des couches sphériques de rayon u centrée à la position \mathbf{r} . L'orbitale 1s est donnée par la formule

$$\rho_{H,\sigma}(r) = \frac{(a^{BR})^3}{8\pi} e^{-a^{BR}r} \quad (1.125)$$

où a^{BR} est un paramètre relié à la charge du noyau, mais pour le modèle BR, il est un paramètre flexible. Suite à l'intégration de la formule 1.124, on obtient l'équation

$$\rho_{x,\sigma}^H(r,u) = \frac{a^{BR}}{16\pi r u} \left[(a^{BR}|r-u|+1)e^{-a^{BR}|r-u|} - (a^{BR}|r+u|+1)e^{-a^{BR}|r+u|} \right] \quad (1.126)$$

où r est la distance entre le point de référence \mathbf{r} et le noyau et u est la distance à partir du point \mathbf{r} . Les auteurs vont se servir de l'équation précédente comme une construction

mathématique et remplacent r par un paramètre b^{BR} ajustable

$$\rho_{x,\sigma}^{BR}(a^{BR}, b^{BR}; u) = \frac{a^{BR}}{16\pi b^{BR}u} \left[(a^{BR}|b^{BR} - u| + 1)e^{-a^{BR}|b^{BR} - u|} - (a^{BR}|b^{BR} + u| + 1)e^{-a^{BR}|b^{BR} + u|} \right]. \quad (1.127)$$

Les paramètres sont ensuite trouvés en reproduisant le développement en série de Taylor au deuxième ordre du trou X exact

$$\rho_{x,\sigma}^{exact}(\mathbf{r}, u) = -\rho_{\sigma}(\mathbf{r}) - Q_{\sigma}(\mathbf{r})u^2 + \dots \quad (1.128)$$

où

$$Q_{\sigma}(\mathbf{r}) = \frac{1}{6} \left[\nabla^2 \rho_{\sigma}(\mathbf{r}) - 2\tau_{\sigma}(\mathbf{r}) + \frac{(\nabla \rho_{\sigma}(\mathbf{r}))^2}{2\rho_{\sigma}(\mathbf{r})} \right]. \quad (1.129)$$

L'autre contrainte pour trouver les paramètres a^{BR} et b^{BR} est le respect de la normalisation du trou X

$$-1 = 4\pi \int_0^{\infty} du u^2 \rho_{x,\sigma}^{BR}(\mathbf{r}, u). \quad (1.130)$$

Le trou BR est particulièrement intéressant d'un point de vue théorique, car il est exact pour les systèmes hydrogénoïdes. Cependant, une de ses limitations est qu'il n'est pas exact pour la limite du gaz homogène, mais il est possible d'ajouter un paramètre pour que le trou BR reproduise la limite du HEG [48].

Du côté des fonctionnelles empiriques, les plus connues sont celles du groupe de Truhlar de l'Université du Minnesota, tel que M06-L [78] et MN15-L [79]. Les Méta-GGA empiriques montrent d'excellents résultats en thermochimie, mais les densités électroniques obtenues sont parfois médiocres [55, 75].

Les Méta-GGA permettent de diminuer davantage la MAE comparativement aux GGA pour le calcul des énergies d'atomisation. Pour la base de données W4-11, la MAE de la fonctionnelle SCAN est de 4 kcal/mol, ce qui est environ la moitié de RevPBE [75] et comparable à la fonctionnelle hybride PBEh. Cependant, PBEh est supérieure à SCAN pour prédire les barrières d'activation.

1.5.5. *Fonctionnelles hybrides.*

Comme le nom l'indique, les fonctionnelles hybrides vont combiner la WFT avec la DFT pour construire une approximation à l'énergie XC en DFT. Le calcul de l'énergie X exacte avec les orbitales est une formule relativement simple pour un déterminant de Slater, cette énergie peut donc être utilisée en KS-DFT. Il pourrait sembler préférable de combiner 100% de l'énergie X exacte avec une fonctionnelle de corrélation provenant d'une GGA ou méta-GGA, néanmoins les résultats seront décevants en pratique [33]. Le problème est que les fonctionnelles d'échange vont inclure en partie de la corrélation statique dans leur énergie X [33]. L'énergie X de HF est exacte et n'inclut pas de corrélation, une fonctionnelle de corrélation conventionnelle semi-locale combinée avec 100% d'échange exact manque donc de corrélation statique. Le trou d'échange exact est parfois délocalisé pour des molécules,

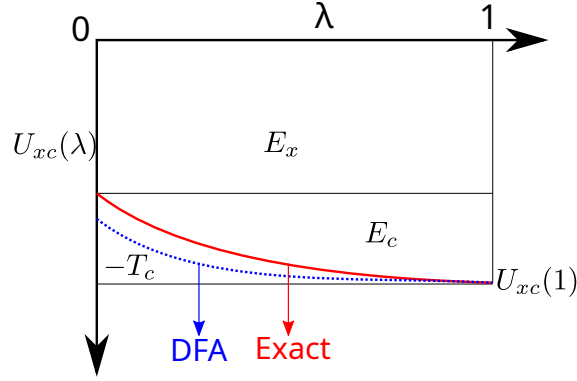


Fig. 1.7. Schéma qui illustre la connexion adiabatique et le fonctionnement des fonctionnelles hybrides. λ est le paramètre de la force de couplage, E_x est l'énergie d'échange, E_c est l'énergie de corrélation, T_c est la contribution de la corrélation à l'énergie cinétique et $U_{xc}(\lambda)$ est l'énergie potentielle d'échange-corrélation qui dépend de λ . DFA signifie *Density Functional Approximation*, donc approximation de la fonctionnelle de la densité.

contrairement au trou XC exact qui est localisé, ce que nous avons discuté à la section 1.4. Ainsi, il faudrait combiner le trou X exact avec une fonctionnelle de corrélation qui elle aussi est délocalisée, afin que la somme du trou X et C soit locale. Le trou d'échange exact (non local) n'est donc pas compatible avec une fonctionnelle de corrélation qui est locale, comme la LDA, les GGA ou les Méta-GGA. Les fonctionnelles d'échange locales vont inclure en partie de la corrélation statique, due à la nature approximative de l'échange. C'est grâce à une annulation d'erreur entre l'échange et la corrélation que les fonctionnelles locales ont de bons résultats. C'est le trou XC total qui est important de reproduire correctement pour calculer E_{xc} et non les trous X et C individuellement.

Ainsi, le chercheur Becke a proposé en 1993 d'utiliser une fraction de l'échange exacte [33]

$$E_{xc}^{hyb} = aE_x^{exact} + (1 - a)E_x^{DFT} + E_c^{DFT}. \quad (1.131)$$

Une représentation schématique de la connexion adiabatique, comme montrée à la figure 1.7, permet de mieux comprendre les fonctionnelles hybrides. L'énergie XC est la somme de l'énergie X (le plus gros rectangle) et C (partie au-dessus de la courbe rouge du deuxième rectangle). Il ne faut pas confondre E_{xc} avec l'énergie potentielle XC (U_{xc}), car cette dernière ne contient pas l'énergie cinétique. L'énergie d'échange est indépendante de λ , ce dernier étant un paramètre qui décrit la force de l'interaction coulombienne entre les électrons et nous avons alors

$$E_{xc}^{\lambda=0} = E_x. \quad (1.132)$$

Pour $\lambda = 1$, l'énergie XC est celle d'un système interagissant et elle n'est pas connue sous une équation mathématique simple. Le deuxième rectangle, en dessous de l'énergie X, a une aire donnée par $U_{xc} - E_x$. L'énergie de corrélation est donc une certaine fraction a de l'aire

de ce rectangle

$$E_c = a(U_{xc} - E_x). \quad (1.133)$$

La fraction a et U_{xc} sont inconnues, ainsi, le principe d'une fonctionnelle hybride est que l'on approxime U_{xc} par une approximation en DFT (E_{xc}^{DFA} , DFA signifie *Density Functional Approximation*) et a est un paramètre à optimiser

$$E_{xc}^{hyb.} = aE_x^{exact} + (1 - a)E_x^{DFA} + E_c^{DFA}. \quad (1.134)$$

Les DFA donnent généralement de bonnes prédictions lorsque $\lambda = 1$, mais c'est à $\lambda = 0$, donc pour l'échange, que ses prédictions sont moins bonnes. L'énergie X d'une DFA est plus basse que celle qui est exacte, il y a donc une partie de l'énergie C qui est incluse, comme schématisée à la figure 1.7. Le paramètre a estime la quantité d'énergie X exacte qui doit être considérée pour s'approcher de la courbe en rouge.

La première fonctionnelle hybride a été développée par Becke [33] et elle se nomme *moitié-moitié*, car il a utilisé la valeur de $a = 0.5$. Dans l'article, Becke a obtenu une MAE de 7 kcal/mol pour le calcul d'énergies d'atomisation d'une base de données, ce qui est une valeur basse pour une fonctionnelle aussi simple. La méthode *moitié-moitié* a donc montré le potentiel des fonctionnelles hybrides en DFT. De nouvelles fonctionnelles hybrides ont ensuite été développées où l'énergie XC de la LDA a été remplacée par l'énergie XC d'une GGA ou Méta-GGA et a est utilisé comme un paramètre empirique. Empiriquement et selon la théorie de perturbation [41], la valeur optimale trouvée pour a se trouve généralement entre 0.2 et 0.3 pour la majorité des systèmes chimiques. Les fonctionnelles hybrides sont extrêmement populaires en chimie, particulièrement la fonctionnelle B3LYP [33, 59, 60], dû à une erreur pour des bases de données en thermochimie qui s'approche de 3 kcal/mol, ce qui est deux fois plus petit que la plupart des Méta-GGA.

Cependant, une limitation importante des fonctionnelles hybrides est que la valeur de a est un paramètre global constant. Par exemple, pour un système unielectronique comme l'hydrogène, la valeur de a doit être d'un en principe, ce qui n'est pas le cas pour la plupart des fonctionnelles hybrides. Il y a un autre type de fonctionnelles qui s'appelle hybrides locales [80, 81] qui vont utiliser $a(\mathbf{r})$, donc la fraction d'échange exact dépend de la position dans l'espace. Ces fonctionnelles seront décrites à la section 1.5.6.

Concernant les résultats des fonctionnelles hybrides, la MAE pour le calcul des énergies d'atomisations de W4-11 de la fonctionnelle B3LYP n'est que de 3 kcal/mol [75]. Le Saint-Graal de la chimie computationnelle est d'obtenir une MAE d'environ 1 kcal/mol, ce qui correspond à l'énergie thermique de la température ambiante lorsque la valeur est arrondie. La MAE assez faible de B3LYP, son temps de calcul raisonnable et son implémentation dans des logiciels lui ont permis de devenir extrêmement populaire en chimie computationnelle. Cependant, B3LYP ne considère pas l'énergie de dispersion et n'est pas adaptée pour les systèmes montrant de la corrélation statique importante comme les métaux de transition

[16, 17]. Ainsi, la recherche fondamentale pour améliorer les fonctionnelles en DFT n'est pas terminée.

1.5.6. Fonctionnelles hybrides locales.

Comme mentionné dans la section précédente, la fraction d'échange exacte optimale (le paramètre a) dépend du système étudié. Les fonctionnelles hybrides locales (LH) [80–86] vont donc corriger ce problème à l'aide d'un paramètre a qui varie selon la position dans l'espace, nous avons ainsi $a_\sigma^{LH}(\mathbf{r})$ et cette quantité se nomme fonction de mélange locale (LMF). La formule générale pour calculer l'énergie XC d'une fonctionnelle hybride locale est la suivante

$$E_{xc}^{LH} = E_x^{exact} + E_c^{DFA} + \sum_{\sigma} \int d\mathbf{r} \rho_{\sigma}(\mathbf{r}) a_{\sigma}^{LH}(\mathbf{r}) (\epsilon_{x,\sigma}^{DFA}(\mathbf{r}) - \epsilon_{x,\sigma}^{exact}(\mathbf{r})) \quad (1.135)$$

et le but du développement de fonctionnelles LH est de trouver un $a_\sigma^{LH}(\mathbf{r})$ qui donne de bons résultats pour diverses régions dans l'espace. Parmi ces LMF, il y a celle de Jaramillo et co-auteurs [82]

$$a_{\sigma}^{LH1}(\mathbf{r}) = 1 - \frac{\tau_{\sigma}^W(\mathbf{r})}{\tau_{\sigma}(\mathbf{r})} \quad (1.136)$$

où on remarque une certaine similarité avec le paramètre $\alpha^{MGG}(\mathbf{r})$ de l'équation 1.123 dans la section 1.5.4 sur les fonctionnelles Méta-GGA. Le paramètre $a_{\sigma}^{LH1}(\mathbf{r})$ permet d'avoir 100% de l'échange exact pour les systèmes unielectroniques comme l'atome d'hydrogène, ce qui corrige l'erreur d'auto-interaction à un électron. Néanmoins, pour le calcul des énergies d'atomisation, les modèles étudiés dans l'article de Jaramillo et co-auteurs montrent des résultats décevants comparativement à B3LYP, avec une MAE environ quatre fois plus grande que B3LYP pour un ensemble de données de 55 molécules [82].

En plus de corriger l'erreur d'auto-interaction, les LMF doivent être adaptés pour inclure de la corrélation statique dans les régions de l'espace appropriées, comme dans la région d'un lien chimique [81, 82]. Les fonctionnelles d'échanges des DFA locales ajoutent en partie de la corrélation statique, ainsi la LMF doit permettre d'avoir une proportion d'échange DFA plus grande dans les régions où la corrélation statique est importante. Jaramillo et co-auteurs [82] ont noté que l'utilisation de la LMF de l'équation 1.136 correspond à une proportion trop grande d'échange exact pour les molécules qu'ils ont étudiées.

Pour corriger les résultats décevants de la fonctionnelle LH de Jaramillo et co-auteurs, une étude par Bahmann et co-auteurs [83] a proposé d'ajouter un paramètre ajustable b

$$a_{\sigma}^{LH2}(\mathbf{r}) = 1 - b \frac{\tau_{\sigma}^W(\mathbf{r})}{\tau_{\sigma}(\mathbf{r})}. \quad (1.137)$$

Ce paramètre a permis d'obtenir une erreur sur le calcul des énergies d'atomisation comparable à B3LYP. Malgré la simplicité de l'équation précédente, les LMF avec un seul paramètre ajustable tel que l'équation 1.137 montrent de bons résultats en thermochimie [81]. L'équation 1.137 a le défaut de ne plus être exacte pour les régions unielectroniques, mais le paramètre, suite à l'ajustement, donne une bonne proportion d'énergie de corrélation statique dans les régions appropriées.

Néanmoins, les LH additionnent les densités d'énergie d'échange d'une DFA avec celles exactes ce qui peut mener à un problème de nom de problème de jauge [81]. Pour les DFA et même pour l'énergie d'échange exacte, la définition de la densité d'énergie n'est pas unique, nous pouvons avoir différentes densités d'énergie qui intègrent pour la même valeur d'énergie aux globales. On peut comprendre le problème de jauge à l'aide d'une fonction $G(\mathbf{r})$ qui a la propriété suivante

$$\int d\mathbf{r}\rho(\mathbf{r})G(\mathbf{r}) = 0. \quad (1.138)$$

Ainsi, lors de l'addition de la fonction précédente avec une densité d'énergie, nous avons

$$\epsilon'_{xc}(\mathbf{r}) = \epsilon_{xc}(\mathbf{r}) + G(\mathbf{r}). \quad (1.139)$$

Donc, les valeurs de $\epsilon'_{xc}(\mathbf{r})$ sont différentes de celle de $\epsilon_{xc}(\mathbf{r})$ localement si $G(\mathbf{r})$ n'est pas une fonction nulle. Néanmoins, globalement, la valeur de l'énergie XC reste identique

$$E'_{xc} = \int d\mathbf{r}\rho(\mathbf{r})\epsilon'_{xc}(\mathbf{r}) = \int d\mathbf{r}\rho(\mathbf{r})\epsilon_{xc}(\mathbf{r}) + \int d\mathbf{r}\rho(\mathbf{r})G(\mathbf{r}) = E_{xc}. \quad (1.140)$$

Lors de la construction de fonctionnelles, des contraintes locales peuvent être ignorées sans avoir un impact majeur sur l'énergie totale, mais un problème de jauge est possible lorsque deux densités d'énergies sont mélangées comme dans un LH. Pour corriger le problème de jauge, on peut modifier l'équation des hybrides locales (équation 1.135) pour obtenir

$$E_x^{LHg} = E_x^{exact} + E_c^{DFA} + \sum_{\sigma} \int d\mathbf{r}\rho_{\sigma}(\mathbf{r})a_{\sigma}^{LH}(\mathbf{r})(\epsilon_x^{DFA}(\mathbf{r}) - \epsilon_x^{exact}(\mathbf{r}) + G_{\sigma}(\mathbf{r})) \quad (1.141)$$

où la fonction $G_{\sigma}(\mathbf{r})$ est la correction de la jauge pour un certain spin [81]. Ces fonctions de correction de jauge sont compliquées et ne seront pas retranscrites ici, mais elles sont construites en respectant diverses contraintes comme l'équation 1.138 et de respecter la limite du HEG [81].

1.5.7. Fonctionnelles hybrides par séparation de portée.

Les fonctionnelles hybrides par séparation de portée (RSH) [84, 86–99] ont une proportion d'énergie d'échange exacte qui s'adapte au système et à la position dans l'espace comme les LH. Néanmoins, la façon d'arriver à cette fin pour les fonctionnelles RSH est assez différente comparativement aux fonctionnelles LH. Les RSH s'intéressent plutôt à séparer l'opérateur de Coulomb en une partie à courte portée (SR) et une autre à longue portée (LR). Les fonctionnelles RSH pour les molécules vont calculer l'échange d'une DFA à SR et l'échange

exact à LR. Les prédictions des DFA à SR sont généralement assez bonnes, mais l'exactitude se détériore à longue portée [91]. Les RSH corrigent en partie ce problème en modifiant l'opérateur de Coulomb selon la formule suivante

$$\frac{1}{r_{12}} = \underbrace{\frac{1 - (\alpha^{RSH} + \beta^{RSH} \operatorname{erf}(\mu^{RSH} r_{12}))}{r_{12}}}_{SR} + \underbrace{\frac{\alpha^{RSH} + \beta^{RSH} \operatorname{erf}(\mu^{RSH} r_{12})}{r_{12}}}_{LR} \quad (1.142)$$

où α^{RSH} , β^{RSH} , μ^{RSH} sont des paramètres et erf est la fonction erreur. L'énergie XC est donnée par la formule suivante

$$E_{xc}^{RSH} = E_c^{DFA} + \alpha^{RSH} E_x^{exact} + \beta^{RSH} E_{x,\mu^{RSH}}^{exact} + (1 - \alpha^{RSH}) E_x^{DFA} - \beta^{RSH} E_{x,\mu^{RSH}}^{DFA}, \quad (1.143)$$

ainsi le paramètre α^{RSH} contrôle la proportion d'énergie d'échange exacte à SR. La somme des paramètres $\alpha^{RSH} + \beta^{RSH}$ indique la quantité d'énergie d'échange exacte à LR dans la fonctionnelle, et lorsque la somme est de 1, nous avons 100% de E_x^{exact} à LR. Finalement, μ^{RSH} contrôle la portée de l'opérateur des RSH.

Lorsqu'il y a 100% d'énergie d'échange exacte à LR, nous avons des fonctionnelles corrigées à longue portée (LC) ce qui permet d'avoir le comportement exact de l'échange à LR. Un exemple d'une fonctionnelle LC est LC- ω PBE [100]. Une fonctionnelle comme CAM-B3LYP [92] utilise environ 20% de E_x^{exact} à SR et 65% à LR.

Pour les solides, le calcul de l'énergie d'échange exacte à longue portée est extrêmement long. Donc, des fonctionnelles RSH comme celle d'Heyd-Scuseria-Ernzerhof (HSE) [90] vont plutôt calculer l'échange exact à courte portée et utiliser une DFA à longue portée. HSE prédit avec une bonne exactitude l'écart énergétique entre la bande de valence et la bande de conduction des semi-conducteurs [101].

Le paramètre μ^{RSH} est trouvé de différentes façons. Par exemple, la fonctionnelle de Brémond et co-auteurs [95] obtient ce paramètre en reproduisant l'énergie de l'hydrogène. Pour la fonctionnelle LC- ω PBE [100], μ^{RSH} est obtenu en minimisant la MAE sur des données de thermochimie.

La majorité des RSH souffre de l'erreur d'auto-interaction à un électron et un paramètre μ^{RSH} global n'est pas adapté pour tous les systèmes. Ces problèmes peuvent être corrigés en utilisant un paramètre de séparation de portée locale, alors nous avons $\mu^{LRSH}(\mathbf{r})$ [93, 98, 102–104]. Ces fonctionnelles se nomment fonctionnelles hybrides par séparation de portée locale (LRSH). Pour un LRSH corrigé avec l'échange exact à LR, où $\alpha^{RSH} = 0$ et $\beta^{RSH} = 1$, l'opérateur coulomb de l'équation 1.142 devient

$$\frac{1}{r_{12}} = \underbrace{\frac{1 - \operatorname{erf}(\mu^{LRSH}(\mathbf{r}_1) r_{12})}{r_{12}}}_{SR} + \underbrace{\frac{\operatorname{erf}(\mu^{LRSH}(\mathbf{r}_1) r_{12})}{r_{12}}}_{LR} \quad (1.144)$$

Lorsque $\mu^{LRSH}(\mathbf{r})$ tend vers 0, c'est la partie à SR de l'équation 1.142 qui domine. À l'opposé, c'est l'opérateur à LR qui est plus grand lorsque le paramètre des LRSH tend vers l'infini. Ainsi, un choix judicieux de $\mu^{LRSH}(\mathbf{r})$ permet de corriger la SIE si ce paramètre tend vers l'infini dans les régions qui correspondent à un système uniélectronique.

Les LRSH de Brütting et co-auteurs [98] montrent d'excellents résultats pour le calcul d'énergies d'atomisation et les barrières d'activation, avec une fonctionnelle ayant des MAE inférieures à 2 kcal/mol pour deux ensembles de données. Leur paramètre local $\mu^{LRSH}(\mathbf{r})$ a été construit pour respecter la limite du HEG, pour que l'énergie d'échange exacte domine à haute densité et pour être exacte pour les systèmes uniélectroniques.

1.5.8. Approximation de la densité pondérée.

La majorité des fonctionnelles populaires ont un souci embarrassant, elles ne prédisent pas correctement l'énergie du système chimique le plus simple: l'atome d'hydrogène. Le problème est que le terme de l'énergie de Coulomb inclut la répulsion d'un électron avec lui-même, ce qui est éliminé exactement par l'énergie d'échange exacte. Pour un système à un électron, nous devons avoir

$$U^{1e}[\rho] + E_x^{1e}[\rho] = 0 = E_c^{1e}[\rho], \quad (1.145)$$

où 1e signifie un système uniélectronique. Une solution simple est d'utiliser 100% de l'échange exact, mais tel que mentionné dans la section des fonctionnelles hybrides (section 1.5.5), il faut alors une fonctionnelle de corrélation non locale ce qui est difficile à construire. Les fonctionnelles LH (section 1.5.6) et LRSH (section 1.5.7) sont d'autres possibilités pour corriger la SIE. Une alternative à l'utilisation de l'échange exacte est de construire une fonctionnelle d'échange non local, qui respecte l'équation suivante lorsqu'il n'y a qu'un seul électron

$$J^{1e}[\rho] = \frac{1}{2} \int \int d\mathbf{r}_1 d\mathbf{r}_2 \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{r_{12}} = -E_x^{1e}[\rho]. \quad (1.146)$$

L'équation précédente montre que si une fonctionnelle d'échange est non locale, donc dépend des positions \mathbf{r}_1 et \mathbf{r}_2 , il est possible de la rendre exacte pour les systèmes uniélectroniques avec $\rho_x^{1e}(\mathbf{r}_1, \mathbf{r}_2) = -\rho(\mathbf{r}_2)$ pour tout \mathbf{r}_1 .

Durant la fin des années 1970, une méthode non locale du nom de l'approximation de la densité pondérée (WDA) a été proposée [105–107]. La WDA a l'avantage non seulement d'être exacte pour les systèmes uniélectroniques, mais aussi de respecter la limite du HEG. Malgré de nombreux développements, les résultats de la WDA pour des systèmes chimiques sont généralement comparables à ceux de la LDA, donc la WDA est peu utilisée en pratique. Il y a encore des chercheurs qui s'intéressent à cette théorie due aux avantages potentiels des méthodes non locales [108–110].

Les approximations dans la WDA reposent sur le trou XC, qui plutôt que d'utiliser la densité électronique, se sert d'une densité pondérée

$$\rho_{xc}[\rho](\mathbf{r},\mathbf{r}') \rightarrow \rho_{xc}^{WDA}[\bar{\rho}](\mathbf{r},\mathbf{r}') \quad (1.147)$$

où la densité pondérée $\bar{\rho}$ est déterminée grâce à la normalisation du trou XC

$$-1 = \int d\mathbf{r}' \rho_{xc}^{WDA}[\bar{\rho}](\mathbf{r},\mathbf{r}'), \quad (1.148)$$

et le trou est donné par l'équation

$$\rho_{xc}^{WDA}[\bar{\rho}](\mathbf{r},\mathbf{r}') = \rho(\mathbf{r}') G^{WDA}(|\mathbf{r} - \mathbf{r}'|, \bar{\rho}) \quad (1.149)$$

où $G^{WDA}(|\mathbf{r} - \mathbf{r}'|, \bar{\rho})$ est une fonction à modéliser. Pour un système uniélectronique, G^{WDA} doit avoir une valeur de -1 partout dans l'espace, puisque $\rho(\mathbf{r}')$ intègre déjà à un seul électron. L'ajustement de la valeur de $\bar{\rho}$ permet à G^{WDA} d'avoir la valeur de -1 pour satisfaire la contrainte de la normalisation du trou XC. Comme la fonction de distribution de paire du HEG est connue et une fonction de distribution est directement reliée au trou XC, l'utilisation de G^{LDA} permet de satisfaire la limite du HEG [109, 110]. L'équation 1.148 est satisfaite lorsque $\rho(\mathbf{r})$ est une constante. Ainsi, la WDA construit un modèle non local qui respecte la limite du HEG et qui est exact pour les systèmes uniélectroniques.

Concernant les études récentes sur la WDA, le groupe d'Ayers [109, 110] s'intéresse à une contrainte possible à satisfaire avec des modèles non locaux, c'est-à-dire la symétrie du trou XC

$$\rho_{xc}(\mathbf{r}_1, \mathbf{r}_2) = \rho_{xc}(\mathbf{r}_2, \mathbf{r}_1). \quad (1.150)$$

Cette propriété du trou XC permet d'obtenir la bonne décroissance asymptotique pour le potentiel XC [109–111]. Pour satisfaire cette contrainte, le groupe d'Ayers a utilisé une équation de Garcia-Gonzalez et co-auteurs [111] où la densité pondérée est remplacée par une quantité non locale qui dépend des positions \mathbf{r}_1 et \mathbf{r}_2 . Cette approche a cependant le désavantage de complexifier considérablement l'obtention de l'équivalent de la densité pondérée. Il devient nécessaire de résoudre des systèmes d'équations non linéaires [109, 110].

Finalement, une méthode non locale inspirée par la WDA et qui a été développée par notre groupe de recherche est le facteur d'échange [108]

$$\rho_{x,\sigma}^{FX}(\mathbf{r}, u) = F_{x,\sigma}^{NL}(\mathbf{r}, u) \rho_{\sigma}(\mathbf{r}, u) \quad (1.151)$$

où F_x^{NL} est le facteur d'échange non local (NL) et $\rho(\mathbf{r}, u)$ est la moyenne sphérique de la densité électronique d'une position \mathbf{r} . La dérivation mathématique de $\rho(\mathbf{r}, u)$ pour des ensembles de bases gaussiens est montrée dans les références [108, 112]. L'utilisation de la moyenne sphérique a l'avantage considérable d'éliminer deux dimensions pour les intégrales, donc il n'est pas nécessaire d'effectuer des intégrations en six dimensions. Néanmoins, contrairement à l'approche du groupe d'Ayers, l'information concernant la symétrie du trou XC (équation

1.150) est perdue avec la moyenne sphérique du trou XC [113]. Le facteur d'échange F_x^{NL} est obtenu à l'aide d'une équation mathématique et des paramètres qui satisfont diverses contraintes comme la normalisation du trou X, la valeur de l'origine du trou X (donc pour $u = 0$) et $F_x = -1$ pour tout u pour les systèmes unielectroniques. Une distinction avec la WDA est que la densité pondérée n'est pas utilisée. Nous allons discuter plus en profondeur du facteur X dans l'article 1.

1.5.9. Problèmes courants des approximations à l'énergie d'échange-corrélation.

Malgré les nombreux succès de la KS-DFT en chimie et en physique, il reste encore de nombreuses limitations des DFA à éliminer. Les limitations courantes des approximations à l'énergie XC en DFT ont été illustrées par Cohen et co-auteurs dans le journal *Science* en 2008 [114] et elles sont encore d'actualité environ 15 ans plus tard. Leur article s'intitule *Insights into Current Limitations of Density Functional Theory*, ce qui signifie: aperçu des limitations courantes en DFT. Les auteurs divisent les échecs des modèles en DFT en deux catégories: les erreurs de délocalisation et de corrélation statique.

La première erreur, l'erreur de délocalisation, est représentée comme étant le fait que les fonctionnelles courantes donnent des erreurs importantes lors du calcul de l'énergie des systèmes ayant des atomes avec un nombre d'électrons fractionnaire. La molécule de H_2^+ étirée est le système la plus simple pour décrire cette limitation, une erreur qui est présente pour des fonctionnelles comme LDA, GGA, Méta-GGA et les hybrides. Néanmoins, la WDA qui a été décrite à la section 1.5.8 est exacte, tout comme HF, pour le H_2^+ . À la figure 1.8, les courbes de dissociation pour HF et pour la fonctionnelle hybride B3LYP sont montrées. HF est une théorie exacte pour les systèmes à un électron comme le H_2^+ , mais B3LYP montre un comportement non physique où l'énergie diminue à partir d'une distance entre les atomes de trois ångströms. Par symétrie, le H_2^+ , à la limite de dissociation, peut être représenté avec des charges fractionnaires: $H^{1/2+}-H^{1/2+}$. Les atomes ont donc chacun la moitié d'un électron. Il a été montré que l'énergie électronique varie linéairement entre N et $N + 1$ électrons [115]. Cependant, la majorité des fonctionnelles montrent une courbe avec une tendance convexe plutôt que linéaire, l'énergie calculée étant trop négative pour des charges fractionnaires. Donc, pour une molécule avec $N + 1/2$ électrons, les fonctionnelles vont sous-estimer l'énergie, comme montré à la figure 1.8 pour B3LYP. HF est une méthode exacte pour les systèmes unielectroniques, mais pour les systèmes multiélectroniques, la courbe montre un comportement concave entre N et $N + 1$. Les auteurs nomment cette limitation de la DFT l'erreur de délocalisation, puisque les fonctionnelles ont la tendance à trop disperser la densité électronique. Une raison qui explique les succès des fonctionnelles hybrides comme B3LYP, est le fait que HF et E_x^{exact} ont un comportement opposé aux LDA/GGA/Méta-GGA. HF et E_x^{exact} souffrent plutôt d'une erreur de localisation pour les

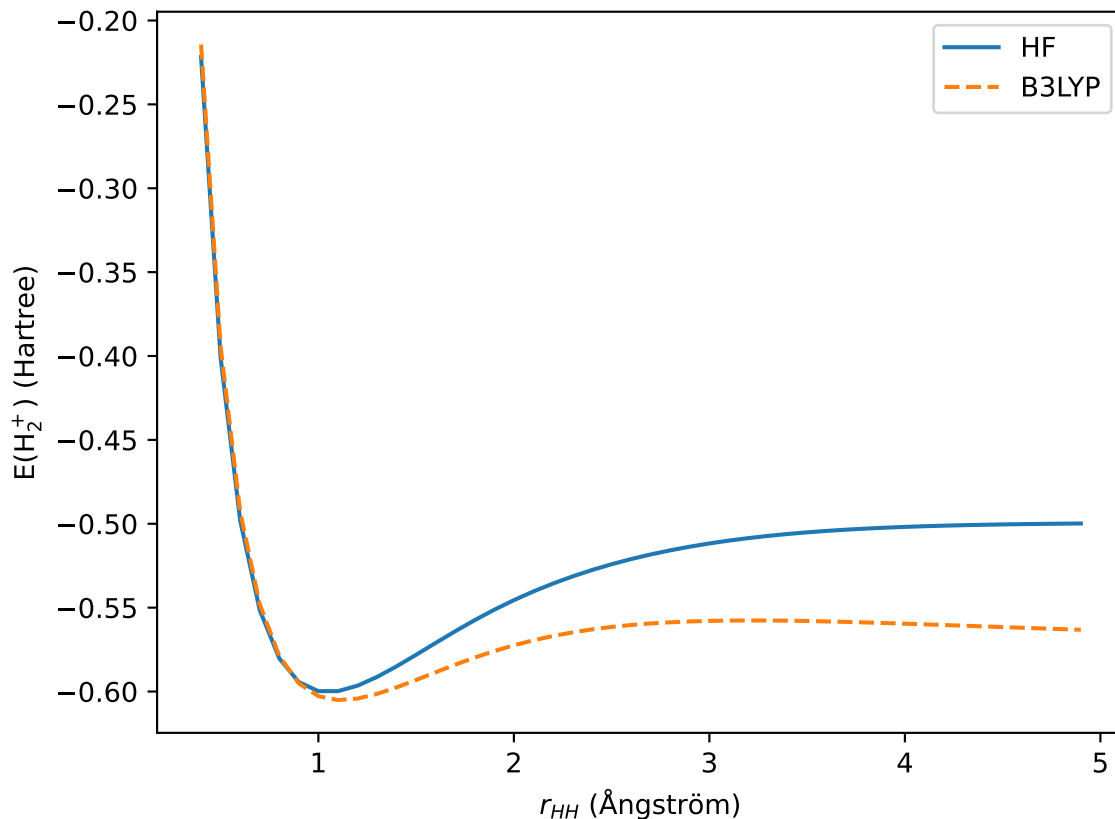


Fig. 1.8. Courbes de dissociation pour la molécule de H_2^+ pour la méthode d’Hartree-Fock ainsi que la fonctionnelle B3LYP. L’ensemble de bases cc-PVTZ est utilisé.

molécules multiélectroniques. Donc, pour une fonctionnelle hybride, il y a une annulation d’erreurs partielle entre la fonctionnelle DFA et l’échange exact.

Ensuite, la deuxième limitation que les auteurs mentionnent est l’erreur associée à la corrélation statique. Nous avons déjà discuté de la molécule de H_2 dissociée et de la nécessité d’utiliser plusieurs déterminants pour décrire ce système (ou d’utiliser UHF/UKS) à la section 1.1.4. Les auteurs de l’article décrivent la corrélation statique avec des spins fractionnaires. Par symétrie, à la limite de dissociation du H_2 , chaque atome d’hydrogène a deux demi-électrons de $1/2$ spin α et $1/2$ spin β , donc l’équivalent d’un électron par atome au total. En principe, l’énergie pour ce type de système de spins fractionnaires doit être la même que celles pour des spins entiers. Par exemple, pour l’atome de Bore, $E(p_x^1 p_y^0 p_z^0) = E(p_x^{1/3} p_y^{1/3} p_z^{1/3}) = E(p_x^0 p_y^1 p_z^0) = E(p_x^0 p_y^0 p_z^1)$ où l’exposant représente le nombre d’électrons présent dans une orbitale, p_x , p_y ou p_z . Par contre, les fonctionnelles en DFT utilisent un seul déterminant de Slater et ont de la difficulté à correctement décrire

ces différents niveaux d'énergie dégénérés, ce qui explique l'erreur importante pour la dissociation du H_2 . Cette erreur est aussi grande pour les métaux de transition, qui ont une structure électronique complexe et qui présentent souvent des niveaux d'énergie dégénérés. La théorie HF n'inclut aucune corrélation statique, alors que les fonctionnelles locales comme LDA/GGA/Méta-GGA vont en avoir une partie due à leurs trous X et C approximatifs qui sont locaux. Les fonctionnelles locales sont donc généralement supérieures, comparativement à celles hybrides, pour calculer les propriétés des métaux de transition [17].

En 2021, des chercheurs chez Deepmind à Londres [116] ont construit une fonctionnelle nommée DM21 à l'aide de l'intelligence artificielle. Ils ont entraîné leur modèle sur des systèmes ayant des électrons fractionnaires. Leurs résultats sont encourageants pour une multitude de systèmes chimiques et il sera intéressant de voir l'impact de la fonctionnelle DM21 dans le futur.

Finalement, une autre limitation aux fonctionnelles courantes en DFT est que les interactions de dispersion ne sont pas représentées [16]. Les interactions de dispersion sont un phénomène à longue portée qui ne peut pas être décrit par des fonctionnelles locales comme les GGA. Pour corriger cette limitation, la façon la plus simple est d'ajouter une correction empirique basée sur l'énergie de dispersion de London [117]

$$E_{disp.} = - \sum_{i,j>i} f_{att.}(R_{ij}) \frac{C_{6ij}}{R_{ij}^6} \quad (1.152)$$

où $f_{att.}(R_{ij})$ est une fonction d'atténuation pour éviter une divergence à la coalescence des noyaux des atomes et C_{6ij} est un coefficient empirique. La correction de dispersion DFT-D3 du groupe de Grimme [118] est basée sur cette approche. Une alternative est la construction d'une fonctionnelle non locale comme l'approche vdW-DF [119]

$$E_{disp.}^{vdW-DF} = \int \int d\mathbf{r}_1 d\mathbf{r}_2 \rho(\mathbf{r}_1) K_C^{nl}(\mathbf{r}_1, \mathbf{r}_2) \rho(\mathbf{r}_2) \quad (1.153)$$

où $K_C^{nl}(\mathbf{r}, \mathbf{r}')$ est une fonction non locale qui est construite à partir d'une fonction de réponse pour décrire les interactions de dispersions. Ces fonctionnelles sont généralement conçues pour des applications en physique du solide, mais il y a tout de même celle de Vydrov et Van Voorhis (VV10) [120] qui a été développée pour des systèmes chimiques. Ces fonctionnelles utilisent souvent des GGA pour calculer l'énergie X et C et l'énergie de Van der Waals est obtenue à partir de l'équation 1.153. Malgré les bonnes performances de VV10, la méthode DFT-D3 est plus rapide à calculer, donc elle demeure la méthode la plus populaire pour modéliser les systèmes ayant des interactions non covalentes importantes [121].

1.6. Ensembles de bases

Pour l'approche KS et HF, les orbitales atomiques sont représentées à l'aide d'une combinaison linéaire de fonctions de base [1–4]

$$\phi_i = \sum_{\mu}^L c_{\mu i} \eta_{\mu} \quad (1.154)$$

où L est le nombre de fonctions de base η_{μ} et $c_{\mu i}$ est un coefficient à optimiser. En chimie computationnelle, le type d'ensembles de bases le plus populaire est le type gaussien (GTO) [1–3], puisque les fonctions gaussiennes sont rapides à intégrer avec un ordinateur. Les orbitales de type Slater (STO) ont une forme exponentielle qui est supérieure aux GTO pour représenter les orbitales hydrogénoïdes, néanmoins les intégrales sont plus longues à calculer d'un point de vue computationnel. La formule générale des GTO est

$$\eta^{GTO}(r) = N x^l y^m z^n e^{-\alpha^{GTO} r^2} \quad (1.155)$$

où N est un facteur de normalisation, α^{GTO} est un paramètre qui contrôle la portée de la fonction gaussienne et l, m, n sont des entiers qui définissent le moment angulaire de l'orbitale. Par exemple, si $l + m + n = 0$, l'orbitale est de symétrie sphérique donc de type s . Malgré l'avantage computationnel, les GTO ont un défaut important, car à la position d'origine (alors pour $r = 0$), la dérivée première d'une GTO est nulle, tandis que les orbitales hydrogénoïdes ont une forme exponentielle et présentent une pointe à l'origine [122]. Une façon d'amoinrir ce désavantage est d'effectuer une combinaison linéaire de fonction gaussienne, comme montré à la figure 1.9. Cette méthode est nommée la contraction de fonctions gaussiennes. Il est ensuite possible de combiner plusieurs fonctions de base de gaussiennes contractées, par exemple la base de Pople 3-21G utilise 3 gaussiennes primitives pour les orbitales des électrons de cœurs et deux fonctions, une composée de 2 gaussiennes et l'autre de 1 gaussienne, pour les électrons de valences. Il est préférable d'utiliser davantage de fonctions pour les électrons de valence, car la chimie se déroule dans ces orbitales. Les électrons de cœurs sont moins influencés par les liaisons chimiques.

Pour représenter des liens chimiques, il faut une certaine flexibilité aux fonctions décrivant les orbitales. Pour la molécule de H_2 , les orbitales individuelles des hydrogènes sont sphériques, mais la symétrie de l'orbitale liante ressemble plutôt à celle d'une orbitale p. Pour améliorer la flexibilité des GTO, des fonctions de polarisations sont ajoutées. Par exemple, la base de Pople 6-31G(d,p) ajoute les orbitales p aux atomes H-He et les orbitales d aux autres atomes.

Finalement, un autre type de gaussienne qui est parfois ajoutée est une fonction gaussienne plus diffuse pour représenter des électrons faiblement liés. Les anions sont un type de système où des fonctions diffuses sont importantes. Dans les bases de Pople, les fonctions diffuses sont représentées comme un "+", par exemple, 6-31+G(d,p).

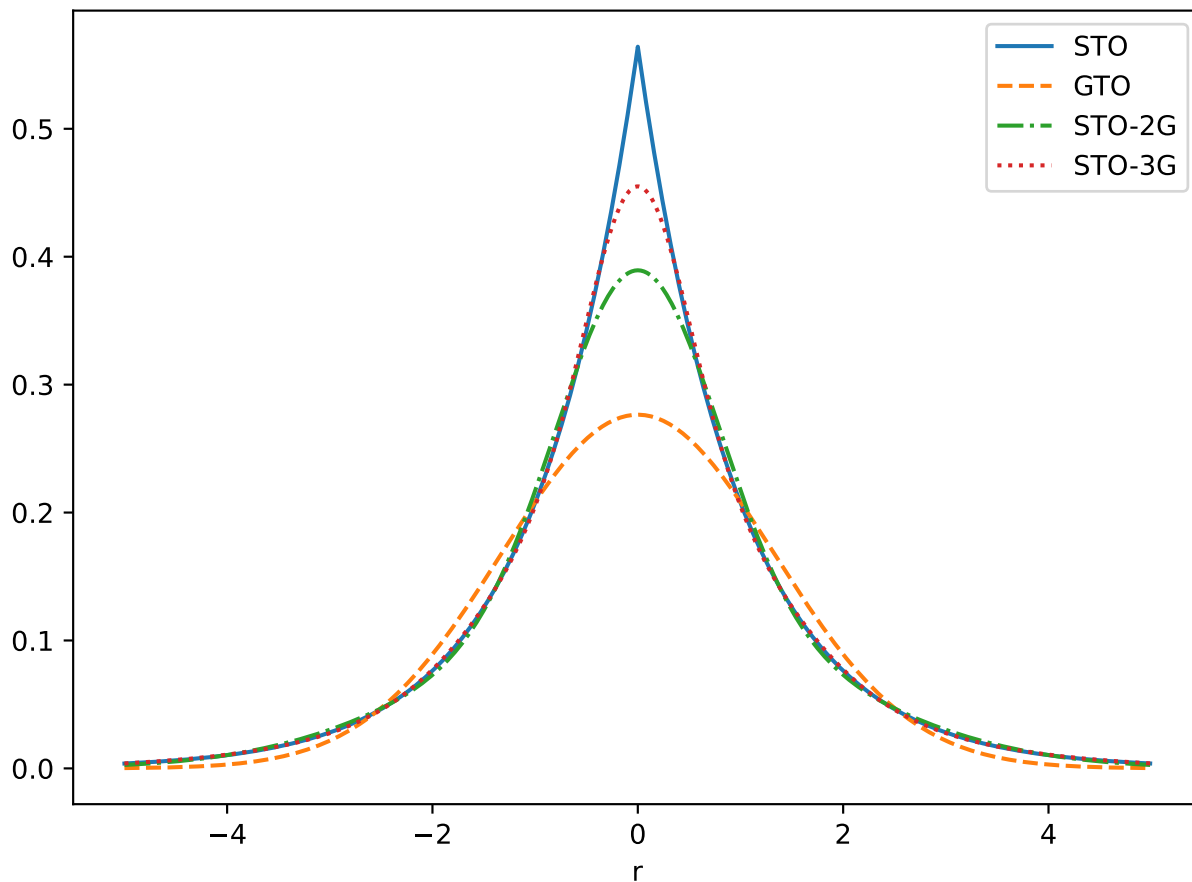


Fig. 1.9. Comparaison entre une orbitale de type Slater (STO), une orbitale de type gaussien (GTO), une somme de deux GTO (STO-2G) et une somme de trois GTO [122]. Les paramètres utilisés dans les GTO proviennent de la référence [122].

Pour les atomes lourds, il faut considérer les effets relativistes et des bases sont construites pour inclure ces effets [8], mais nous n'en discuterons pas davantage dans cette thèse. Nous étudions surtout des petites molécules organiques simples afin de tester nos fonctionnelles, donc les effets relativistes sont négligeables.

2. Apprentissage automatique

L'apprentissage automatique (ML) est une sous-branche du domaine de l'intelligence artificielle (IA). Le ML se sert des statistiques, de données massives et des ordinateurs pour accomplir diverses tâches. À l'aide des données, le ML construit des modèles qui trouvent, en d'autres mots "apprennent", des relations et des tendances qui sont présentes dans ces données. Les méthodes en ML varient beaucoup en complexité, par exemple, la plus simple est la régression linéaire. À l'autre extrême, l'apprentissage profond accomplit des tâches complexes, comme de reconnaître des animaux dans une image. La progression de la puissance de calcul des ordinateurs, l'augmentation grandissante de la quantité de données disponibles et de nouveaux algorithmes plus efficaces ont permis au ML de prendre une place importante dans tous les domaines scientifiques [123].

Il y a quatre types de méthodes principales en ML [124], les apprentissages:

- Supervisé: c'est-à-dire qu'un modèle est entraîné à partir de données étiquetées, donc il faut fournir des valeurs que le ML doit reproduire avec une certaine précision. Cette approche peut classifier des objets (valeurs discrètes) ou effectuer des régressions (valeurs continues). La section 2.1 détaillera davantage l'apprentissage supervisé.
- Non supervisé: ce type de ML trouve des relations dans les données d'entraînement sans qu'il y ait d'étiquettes ou de réponse connue. Cette approche regroupe des données selon des caractéristiques communes entre-elles (partitionnement de données), trouve des relations entre les données ou diminue la dimensionnalité du problème. La section 2.2 détaillera davantage l'apprentissage non supervisé.
- Semi-supervisé: Comme son nom l'indique, cette méthode va combiner l'apprentissage supervisé et non supervisé. Certaines données sont étiquetées, alors que d'autres ne le sont pas. Cette méthode sera détaillée dans la section 2.2.
- Renforcement: L'apprentissage par renforcement est basé sur un agent qui interagit avec son environnement. L'agent peut recevoir une réponse positive ou négative de l'environnement suite à une action. Cette méthode sera détaillée dans la section 2.3

Ces méthodes de ML permettent d'accomplir des tâches variées, telles que de reconnaître des images, des textes et même de pouvoir en générer. L'apprentissage profond [125] est probablement l'approche d'intelligence artificielle dont on entend le plus souvent parler dans les médias en 2023. Par exemple, le robot de discussion en ligne ChatGPT de la compagnie OpenAI risque d'avoir une influence considérable en enseignement [126] et en écriture d'articles scientifiques [127]. Selon nos tests personnels du début de l'année 2023, ChatGPT permet même d'écrire des codes Python pour faire des calculs selon le modèle Hartree-Fock. Il est cependant important de vérifier les résultats, car il est fréquent que des erreurs se glissent dans le code. Toutefois, ChatGPT montre tout de même le grand potentiel de l'intelligence artificielle en science.

Malgré la quantité phénoménale d'applications en tout genre du ML, il existe tout de même plusieurs limitations importantes à cette branche de l'IA. Premièrement, il faut générer des données de haute qualité pour entraîner de bons modèles. Si les données d'entraînements comportent trop d'erreurs, il est évident que le ML ne produira pas des prédictions raisonnables. Il y a une phrase importante en anglais qui explique bien ce phénomène : *Garbage in, garbage out* (GIGO), ce qui signifie déchets en entrées, déchets à la sortie. De plus, les modèles en ML sont souvent composés de milliers de paramètres ajustables. Ils sont donc très difficiles à interpréter et ne permettent pas une compréhension simple derrière les prédictions. D'ailleurs, les techniques de ML sont généralement très inefficaces pour faire des prédictions en dehors des caractéristiques des données qui ont été utilisées lors de l'entraînement. La référence [128] montre qu'un réseau de neurones est capable de reproduire une fonction sinusoïdale à l'intérieur de l'intervalle de valeurs qui a été utilisé lors de l'entraînement, mais à l'extérieur de cet intervalle, les prédictions du réseau de neurones sont complètement erronées.

Afin d'obtenir un modèle ML qui est le plus généralisable possible, un concept essentiel à comprendre est le compromis entre le biais et la variance. Ce compromis est expliqué à la figure 1.10. Le biais est associé à l'erreur obtenue lors de l'apprentissage sur des données d'entraînements. Un biais élevé signifie donc un sous-apprentissage, car le modèle de ML n'a pas correctement appris les tendances dans les données. La variance est la sensibilité du modèle aux fluctuations aléatoires dans un ensemble de données. Si un modèle reproduit du bruit dans les données, la variance est élevée et ça signifie un surapprentissage, voir la figure 1.11 pour un exemple qui illustre ce phénomène. Pour s'assurer qu'un modèle est le plus généralisable possible, on peut utiliser un ensemble de données pour l'entraînement, un autre pour valider le modèle et un dernier pour tester le modèle obtenu suite à l'apprentissage [129]. Les données d'entraînement sont utilisées pour optimiser les paramètres du modèle, par exemple, afin de trouver les paramètres a et b d'une équation linéaire de type $ax + b$. Les données de validations vérifient que le modèle ne souffre pas de surapprentissage et ajustent les hyperparamètres, c'est-à-dire des paramètres qui contrôlent l'apprentissage, comme des variables associées à l'algorithme d'optimisation. Finalement, les données tests servent à tester le modèle sur un grand nombre de données qui n'ont pas été utilisées lors de l'entraînement et la validation.

Pour en savoir plus sur les sujets de l'apprentissage automatique et des réseaux de neurones, les livres *Deep Learning with PyTorch* [129] et *Deep Learning* [130] sont recommandés.

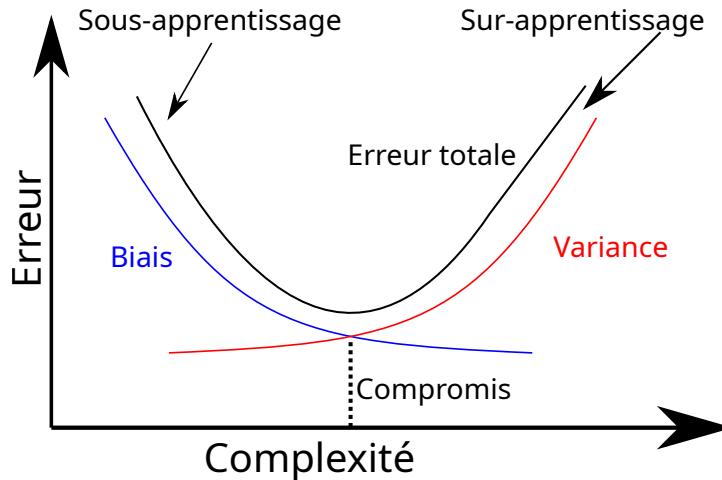


Fig. 1.10. Le compromis entre le biais et la variance. Le biais est associé à l'erreur lors de l'apprentissage sur les données d'entraînement, tandis que la variance est la sensibilité du modèle aux fluctuations aléatoires dans un ensemble de données. Un modèle trop simple pour un problème mène à une erreur importante causée par un biais élevé, ce qui correspond à un sous-apprentissage. À l'opposé, un modèle trop complexe pour un problème pourrait souffrir de surapprentissage et avoir une variance élevée. Un modèle qui a surappris des tendances dans les données n'est donc pas généralisable, malgré une erreur faible lors de l'entraînement.

2.1. Apprentissage supervisé

L'apprentissage supervisé construit une fonction approximative grâce à des données d'entraînement de type (x, y) où x sont des valeurs d'entrées et y sont des données étiquetées à reproduire. En d'autres mots, il faut fournir des exemples de données d'entrées et de sorties pour que le modèle puisse "apprendre" la relation entre les deux. La réponse à notre problème est en partie connue, mais il faut trouver le chemin pour arriver à cette réponse. Ainsi, une fois le modèle entraîné, une prédiction y^* est obtenue, alors $y^* = f(x^*)$ où f est une fonction d'une méthode ML comme un réseau de neurones et x^* est une entrée différente de celle utilisée lors de l'entraînement du modèle. Le choix de la méthode ML va dépendre du problème à résoudre. Une régression linéaire ne fonctionne pas correctement pour reproduire une relation non linéaire et un réseau de neurones n'est pas nécessaire pour résoudre un problème linéaire.

L'entraînement d'un modèle ML est effectué en optimisant les paramètres grâce à une fonction de coût $C(h(x, w), y)$, où l'argument $h(x, w)$ est une fonction constituée de paramètres w qui est ajustée pour diminuer l'erreur de $h(x, w)$ par rapport aux étiquettes y . Un exemple de fonction $h(x, w)$ est une fonction linéaire ($y^* = w_1 x + w_2$) où w_1 et w_2 sont les paramètres

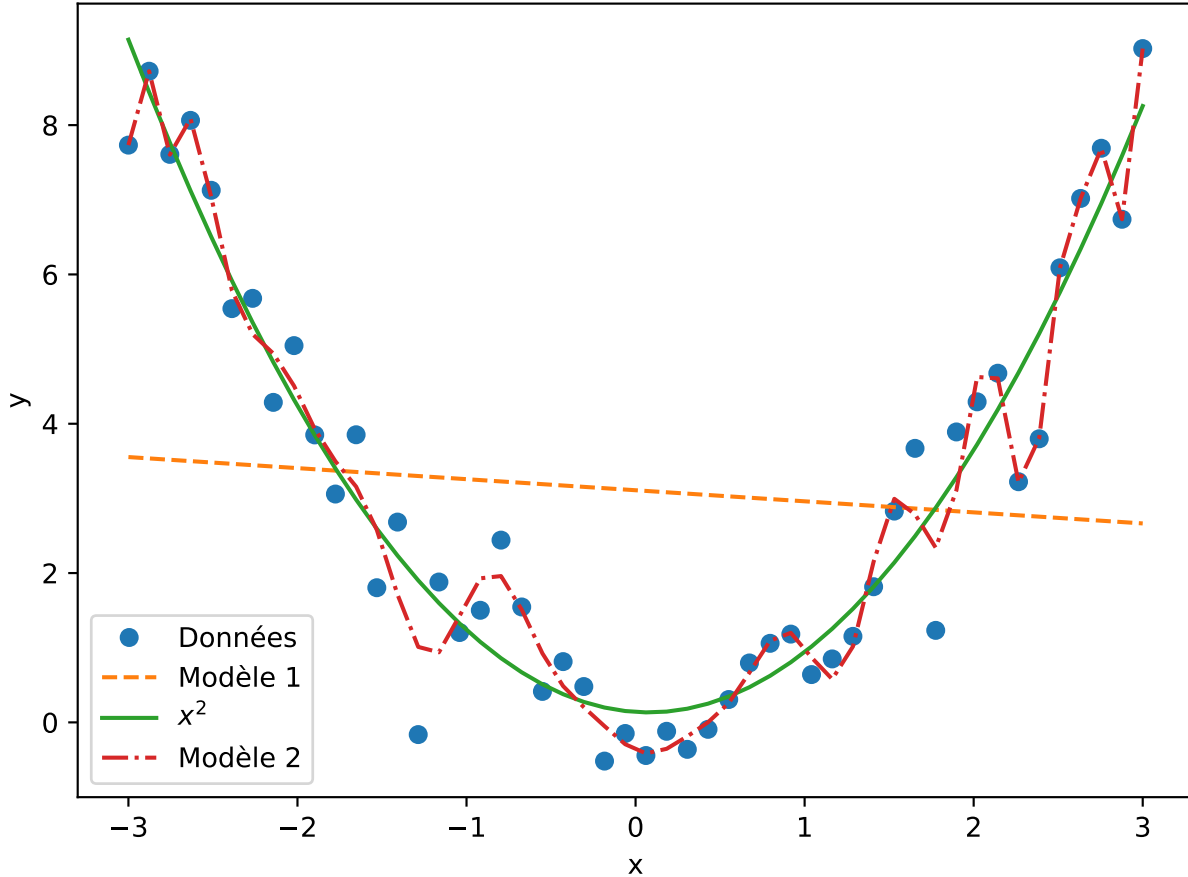


Fig. 1.11. Les données représentent une fonction quadratique avec un bruit aléatoire de type gaussien. Le modèle 1 est une régression linéaire, tandis que le modèle 2 est un polynôme de degré 30. Les paramètres des modèles ont été obtenus à l'aide d'un ajustement par moindres carrés. Le modèle 1 souffre de sous-apprentissage tandis que le modèle 2 a une variance élevée, donc il a surappris les tendances dans les données.

à optimiser. Une fonction de coût très populaire est l'erreur moyenne au carré (MSD)

$$C^{MSD}(h(x,w),y) = \frac{1}{N} \sum_i (y_i^* - y_i)^2, \quad (2.1)$$

ce qui estime le carré de la distance entre les prédictions (y_i^*) de la fonction hypothèse h et les données étiquetées (y). Cette distance doit être minimisée pour obtenir les paramètres optimaux. Cette fonction de coût est positive et sera nulle si le modèle prédit parfaitement les valeurs. Cependant, comme montré aux figures 1.10 et 1.11, ce n'est pas parce qu'un modèle ML a une erreur nulle sur les données d'entraînement qu'il est généralisable pour d'autres valeurs d'entrées qui sont différentes de celles utilisées lors d'entraînement.

L'apprentissage supervisé peut servir pour effectuer des régressions (données de sortie continues) ou classifier des objets (données de sortie discrètes). La figure 1.12 montre un

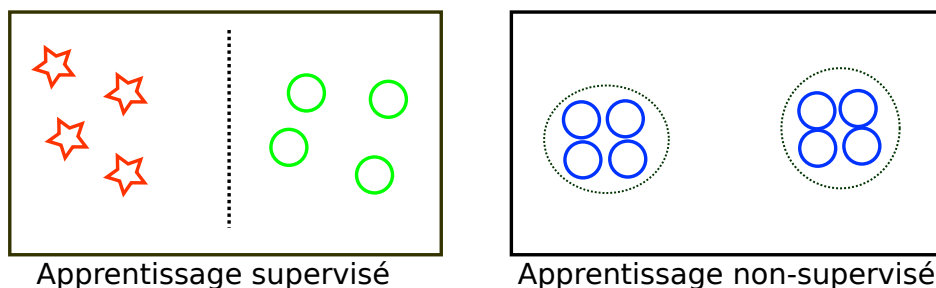


Fig. 1.12. À gauche, un exemple de classification à l'aide de l'apprentissage supervisé. Une méthode ML est entraînée à partir de données qui sont représentées par des ronds et des étoiles, pour ensuite classifier des entrées futures afin de savoir si elles sont un rond ou une étoile. À droite, un exemple d'apprentissage non supervisé pour partitionner des données. Pour l'apprentissage non supervisé, c'est au ML de séparer les données, l'information concernant si les entrées sont des étoiles ou des ronds n'est pas présente.

exemple de classification par apprentissage supervisé. En chimie, le ML pourrait être utilisé pour prédire l'énergie en fonction de la distance entre deux atomes. Près de la longueur d'équilibre, selon l'approximation harmonique, l'énergie doit varier selon une équation quadratique en fonction de la distance de la longueur de lien à l'équilibre. Les réseaux de neurones sont d'ailleurs utilisés pour construire des surfaces d'énergie potentielle de molécules [131]. Concernant la classification, des chercheurs se sont servis du ML afin de construire un modèle qui identifie des molécules organiques à partir de leur indice de réfraction [132].

Un exemple d'apprentissage supervisé, qui est utilisé dans la vie de tous les jours, est de classifier si un courriel est du pourriel ou non [133]. À chaque courriel (l'entrée x), on associe une étiquette "courriel" ou "pourriel" (la donnée étiquetée y). Ensuite, on choisit un algorithme de ML approprié [134], puis pour donner suite à l'entraînement, on peut lui donner une nouvelle entrée x^* et l'algorithme va tenter de prédire si le courriel est du pourriel ou non (la sortie y^*).

2.2. Apprentissage non supervisé

La définition de l'apprentissage non supervisé est lorsqu'un modèle ML n'utilise pas de données étiquetées (donc les données nommées y dans la section 2.1) afin d'optimiser ses paramètres. Ce type de ML va plutôt s'intéresser à trouver des informations, des corrélations ou des structures qui sont présentes dans l'ensemble de données. Par exemple, l'apprentissage non supervisé sert à regrouper des données qui montrent des caractéristiques communes, voir la figure 1.12 pour une illustration. Lorsqu'il y a des relations dans les données, ça peut permettre au ML de réduire la dimensionnalité du problème.

Les modèles utilisant l'apprentissage non supervisé sont généralement plus difficiles à optimiser et à utiliser que l'apprentissage supervisé, car il n'y a pas d'étiquettes qui permettent

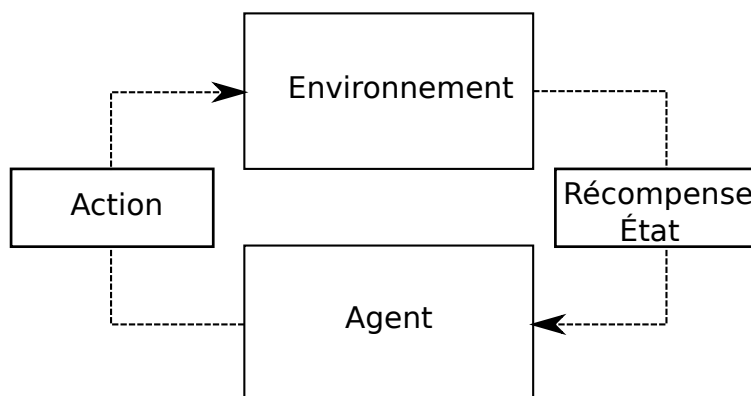


Fig. 1.13. Schéma qui explique le principe de l'apprentissage par renforcement. Un agent fait une action pour interagir avec l'environnement. Ce dernier va retourner une récompense ou une punition. L'état représente la situation actuelle de l'agent dans l'environnement.

de vérifier les résultats. Concernant l'exemple de la figure 1.12, il pourrait être possible que l'algorithme ne sépare pas correctement les deux groupes, l'intuition d'un expert est ainsi essentielle. Un exemple d'application du ML non supervisé en chimie est une étude qui a utilisé le ML pour analyser le mécanisme d'autoassemblage de particules [135].

De plus, il est possible de combiner l'apprentissage supervisé avec celui non supervisé, ce qui se nomme l'apprentissage semi-supervisé. Un groupe de recherche en chimie [136] s'est servi de l'apprentissage semi-supervisé pour classifier les systèmes chimiques en fonction de la quantité de corrélation statique présente dans la molécule. Le but de leur étude est de déterminer si une méthode qui utilise plusieurs déterminants de Slater, comme MCSCF, est nécessaire pour correctement représenter la structure électronique du système. Le groupe a étiqueté quelques données montrant une corrélation statique plus forte que les autres, ce qui leur a permis d'obtenir des résultats supérieurs à l'apprentissage non supervisé en s'assurant que les molécules montrant des comportements moins fréquents soient correctement considérées dans leur modèle ML.

2.3. *Apprentissage par renforcement*

Un autre type de ML est l'apprentissage par renforcement, le schéma de la figure 1.13 donne une explication visuelle. L'apprentissage par renforcement permet au modèle d'apprendre, grâce à une réponse positive ou négative de l'environnement suite à une action de l'agent. Ce type de ML est particulièrement populaire pour les applications dans le domaine des jeux vidéos. Comme exemple, l'apprentissage par renforcement a créé une intelligence artificielle qui est capable de jouer au jeu classique Pong [137]. L'agent est la palette qui repousse la balle derrière le but de son adversaire. L'action que l'agent peut accomplir dans son environnement est de se déplacer vers le haut ou vers le bas. Si l'agent va dans la bonne

direction, il va avoir une récompense en repoussant la balle derrière le but adverse. Sinon, il a une punition lorsque la balle se retrouve derrière lui. L'état de l'agent dans le jeu Pong est sa position actuelle.

En chimie, l'apprentissage par renforcement sert, par exemple, à améliorer les algorithmes d'optimisation des géométries moléculaires [138]. Les auteurs de l'article ont utilisé le ML pour assister les algorithmes d'optimisation à trouver la meilleure direction que les atomes doivent suivre afin de minimiser l'énergie et d'obtenir la structure d'équilibre du système.

2.4. Réseaux de neurones

Maintenant que nous avons introduit les différents types de ML, nous allons discuter de la méthode ML que nous avons utilisée dans cette thèse. Il y a plusieurs méthodes en ML, comme la régression ridge à noyau, l'algorithme des K plus proches voisins, les arbres de décisions, etc [124]. Due à la possibilité de construire sa propre fonction de coût, ce sont les réseaux de neurones artificiels (NN) [125, 130], en particulier les perceptrons multicouches (MLP), que nous avons utilisés lors de nos projets de cette thèse. Il y a d'autres types de NN comme les NN convolutifs et les NN récurrents [124], cependant nous n'en discuterons pas dans cette section. Ce sont des méthodes plus complexes que les MLP et leur utilisation n'a pas été nécessaire pour accomplir les projets de cette thèse. Dans les projets des articles 2 et 3, nous allons nous servir des MLP pour faire de l'apprentissage supervisé et non supervisé sur des données continues. Nous allons maintenant décrire l'architecture des MLP, ensuite le processus utilisé pour optimiser leurs paramètres ajustables.

Tel que le nom des NN le suggère, les principes mathématiques derrière leur conception sont inspirés par la biologie du cerveau humain. Un des premiers modèles NN se nomme perceptron et il est constitué d'un neurone

$$f(\mathbf{x}) = \text{sign}\left(\sum_{i=1}^N w_i x_i - b\right) \quad (2.2)$$

où w_i représentent les poids et b est le biais, ces deux variables étant des paramètres ajustables. La fonction *sign* retourne une valeur de 1 si la valeur à l'intérieur du neurone est plus grande que zéro et la valeur de -1 sinon. Les poids déterminent si l'entrée a une influence importante ou non sur la fonction, tandis que le biais déplace la fonction. Les fonctions que le perceptron peut approximer sont très limitées, il ne peut pas approximer la fonction "OU exclusif" (XOR) par exemple. Néanmoins, la forte progression de la puissance de calcul des ordinateurs, ainsi que des algorithmes plus efficaces ont permis l'avènement des perceptrons multicouches (MLP)

$$f(\mathbf{x}) = \sigma^{NN}(\mathbf{W}_n(\dots\sigma^{NN}(\mathbf{W}_2(\sigma^{NN}(\mathbf{W}_1\mathbf{x} + \mathbf{b}_1) + \mathbf{b}_2))\dots + \mathbf{b}_n)) \quad (2.3)$$

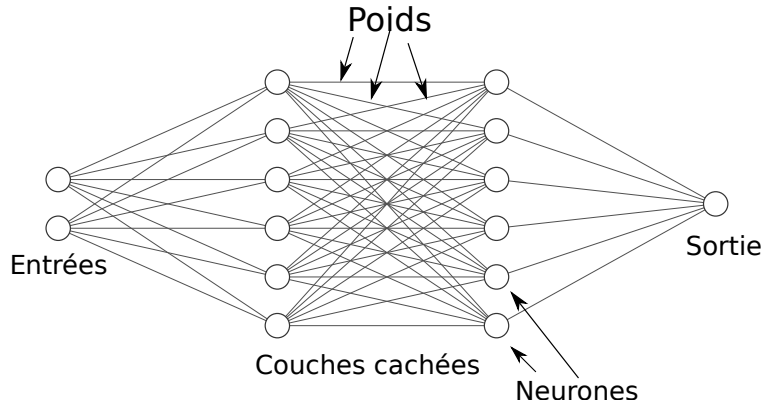


Fig. 1.14. Exemple schématique d'un réseau de neurones artificiels avec deux variables d'entrées, deux couches cachées de six neurones chacune, ainsi qu'une variable de sortie. Les lignes dans ce schéma représentent les poids. Ce schéma a été construit à l'aide de l'outil de la référence [140].

où σ^{NN} est une fonction dite d'activation, \mathbf{b} est un vecteur de biais et \mathbf{W} est une matrice de poids. Les poids et les biais sont les paramètres du MLP à optimiser lors de l'entraînement, donc lors de la minimisation d'une fonction de coût C . Un exemple de l'architecture d'un perceptron multicouche est montré à la figure 1.14. Il existe une grande quantité de fonctions d'activation σ^{NN} [129] et le choix est important. Il y a, par exemple, les fonctions tanh ou sigmoïde, qui sont inspirées par l'activation d'un neurone dans le cerveau, puisqu'elles vont saturer à une valeur qui correspond à un neurone activé. Selon le théorème de l'approximation universelle, les NN ont la capacité de reproduire toutes les fonctions continues [139], ce qui est un très grand avantage. Cependant, une propriété essentielle que la fonction d'activation σ^{NN} doit satisfaire est qu'elle doit être une fonction non linéaire, sinon le NN sera limité à l'approximation de fonctions linéaires. La fonction d'activation doit aussi être différentiable, afin de pouvoir calculer le gradient de σ^{NN} qui est nécessaire lors de l'optimisation des paramètres. Une discontinuité dans la dérivée à un point en particulier n'est pas un problème en pratique, par exemple, la populaire fonction d'activation ReLU ($\sigma^{ReLU}(x) = \max(0, x)$) montre une discontinuité dans sa dérivée à l'origine. Le choix de la fonction d'activation est assez difficile, il est préférable d'en essayer plusieurs et d'étudier les résultats obtenus par le MLP. Certaines peuvent permettre à la fonction de coût d'être plus facile à converger lors de l'optimisation ou d'autres peuvent donner des résultats supérieurs.

Dorénavant dans ce texte, l'acronyme NN va faire référence aux MLP, à moins d'une mention du contraire.

2.4.1. Algorithme d'optimisation des paramètres d'un réseau de neurones.

Afin d'obtenir les poids et les biais optimaux, il faut minimiser une fonction de coût

$$\min_{\mathbf{W}, \mathbf{b}} C(\mathbf{x}, \mathbf{W}, \mathbf{b}) \quad (2.4)$$

en optimisation les poids et les biais. Le choix de la fonction de coût C est extrêmement flexible, les MLP peuvent effectuer de l'apprentissage supervisé tout comme celui non supervisé.

Afin d'obtenir les dérivées de la fonction de coût qui donne les paramètres optimaux, l'algorithme de la rétropropagation du gradient est le plus couramment utilisé [141, 142]. Pour une fonction de coût C , le but de cet algorithme est d'optimiser les poids et les biais à l'aide des dérivées partielles $\partial C/\partial w$ et $\partial C/\partial b$. Pour trouver ces dérivées, l'algorithme de la rétropropagation du gradient se sert de la règle de chaîne et de la différentiation automatique. La différentiation automatique retourne les dérivées d'une fonction à l'aide de la règle de chaîne et des opérations mathématiques élémentaires effectuées par la fonction. Les logiciels de ML comme PyTorch [143] implémentent la différentiation automatique, ce qui est une méthode très puissante pour obtenir les dérivées d'une fonction.

Une fois que les dérivées de la fonction de coût en fonction des poids et des biais sont obtenues, nous allons maintenant décrire quelques algorithmes que nous avons utilisés lors des projets de cette thèse pour optimiser les paramètres. L'algorithme le plus simple pour optimiser les paramètres d'un NN se nomme la descente du gradient

$$w_{t+1} = w_t - \alpha \frac{\partial C}{\partial w_t} \quad (2.5)$$

où t est le temps (ici le temps est relié aux nombres d'itérations), α est un hyperparamètre (les hyperparamètres contrôlent l'optimisation) appelé le "taux d'apprentissage" et C est une fonction de coût. L'hyperparamètre α est très important et il est difficile à estimer en pratique. Si l'hyperparamètre α est trop petit, l'optimisation est extrêmement lente. Si le α est trop grand, le changement du paramètre w est possiblement trop brusque et il est difficile de converger vers le minimum de la fonction. L'algorithme de la descente du gradient est donc très simple, mais ce n'est pas le plus robuste pour optimiser les paramètres.

Un algorithme alternatif permet à l'hyperparamètre α de s'adapter au problème lors de l'optimisation. De plus, α est différent pour chaque poids. Cet algorithme est nommé l'algorithme Adam [144] (Adam signifie *Adaptive moment estimation*). Cet algorithme est une combinaison de deux algorithmes : l'algorithme de la descente du gradient avec momentum et de la propagation de la valeur quadratique moyenne (RMSProp).

L'algorithme de la descente du gradient avec momentum optimise les poids à l'aide d'une moyenne mobile du gradient

$$w_{t+1} = w_t + \alpha m_t \quad (2.6)$$

où

$$m_t = \beta m_{t-1} + (1 - \beta_1) \frac{\partial C}{\partial w_t} \quad (2.7)$$

et où β est un hyperparamètre dont la valeur contrôle l'influence de l'itération précédente. Généralement, la valeur de $\beta_1 = 0,9$ est choisie [130]. L'utilisation du momentum m_t diminue les oscillations du gradient lors de l'optimisation.

L'algorithme RMSProp, quant à lui, va permettre d'utiliser un α qui évolue lors de l'optimisation plutôt que de demeurer constant. Le poids à l'itération $t + 1$ est donné par la formule

$$w_{t+1} = w_t - \frac{\alpha}{\sqrt{v_t + \epsilon}} \frac{\partial C}{\partial w_t} \quad (2.8)$$

où ϵ est une petite valeur comme 10^{-8} afin d'éviter une division par 0 et

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) \left(\frac{\partial C}{\partial w_t} \right)^2 \quad (2.9)$$

où β_2 est une constante qui est généralement égale à 0,9. Le paramètre α est divisé par la racine de v_t , ce qui permet indirectement au α d'être adaptatif à la taille du gradient du coût. Chaque paramètre w_t va avoir un taux d'apprentissage différent des autres poids, ce qui facilite l'optimisation.

Comme mentionné précédemment, l'algorithme Adam combine les avantages de la descente du gradient avec momentum et l'algorithme RMSProp. On se retrouve avec

$$w_{t+1} = w_t - \alpha \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}} \quad (2.10)$$

où

$$\hat{m}_t = \frac{m_t}{1 - \beta_1} \quad (2.11)$$

et

$$\hat{v}_t = \frac{v_t}{1 - \beta_2}. \quad (2.12)$$

Cet algorithme est très fréquemment utilisé pour optimiser les paramètres des NN, car il est efficace et ne consomme pas beaucoup de mémoire.

Finalement, le dernier algorithme qui sera discuté brièvement dans ce chapitre est l'algorithme de Broyden, Fletcher, Goldfarb et Shanno (BFGS) [145–148]. Cet algorithme est de deuxième ordre, car il utilise aussi la deuxième dérivée de la fonction de coût par rapport aux paramètres. Mais il est considéré comme étant "Quasi-Newton", puisque la deuxième dérivée est approximée. L'avantage des algorithmes de deuxième ordre est que la deuxième dérivée donne de l'information concernant le paramètre α optimal à utiliser. La formule de la méthode de Newton est la suivante

$$w_{t+1} = w_t + \frac{C'(w_t)}{C''(w_t)} \quad (2.13)$$

où $C'(w_t)$ représente la première dérivée de la fonction de coût par rapport au poids w_t et $C''(w_t)$ est la dérivée de second ordre. On remarque d'ailleurs que le paramètre α de l'algorithme de la descente du gradient n'est plus présent, il est remplacé par la dérivée de second ordre. La dérivée de second ordre est souvent longue à calculer et le but des méthodes dites Quasi-Newton est de l'approximer. Les équations mathématiques des algorithmes d'optimisation de second ordre sont assez compliquées et ne seront pas retranscrites ici.

2.5. Apprentissage automatique appliqué à la théorie de la fonctionnelle de la densité

Comme beaucoup de domaines en science, l'apprentissage machine (ML) gagne en popularité chez les théoriciens en DFT, comme le démontrent ces nombreux articles de revue qui ont été publiés entre 2020 et 2023 [149–153]. Le ML trouve des relations complexes qui sont cachées dans les données, ce qui peut aider à obtenir de nouvelles approximations à l'énergie électronique en DFT.

L'un des premiers articles concernant l'utilisation du ML en DFT a été publié en 2012 par Snyder et co-auteurs [154]. Une aire de recherche en DFT, alternative à l'approche Kohn-Sham, vise à éviter l'utilisation des orbitales en construisant des approximations à l'énergie cinétique. Cette méthode se nomme *Orbital-Free DFT* et le ML est une approche qui pourrait améliorer les fonctionnelles d'énergie cinétique [150]. Les auteurs ont utilisé la méthode ML qui se nomme la régression ridge à noyau, pour approximer l'énergie cinétique d'un système simple constitué de fermions non interagissant dans une boîte en une dimension. Le ML évite la résolution de l'équation de Schrödinger une fois que le modèle est entraîné. Pour le calcul de l'énergie cinétique, leur approche de ML leur a permis d'obtenir une erreur inférieure à 1 kcal/mol. Cette étude a donc montré le potentiel du ML en DFT.

Concernant l'utilisation du ML pour construire des approximations à l'énergie XC selon l'approche Kohn-Sham, l'étude pionnière qui a influencé nos travaux est celle de Nagai et co-auteurs qui a été publiée en 2020 [155]. Leur étude est particulièrement motivante dans le domaine du développement de fonctionnelles XC, puisqu'ils ont montré qu'à l'aide d'une fonctionnelle construite par un NN, les résultats obtenus sont comparables ou mieux que les approximations courantes en DFT. Le plus surprenant est que leur ensemble de données d'entraînement est composé de seulement trois molécules (H_2O , NO , NH_3). L'information "cachée" qui se trouve dans ces molécules permet ainsi de construire un modèle généralisable pour des molécules organiques simples. Le titre de leur article se nomme d'ailleurs *Completing Density Functional Theory by Machine Learning Hidden Messages from Molecules*.

L'équation qu'ils ont utilisée pour calculer la densité d'énergie XC est la suivante

$$\epsilon_{xc}^{NN}(\mathbf{r}) = \epsilon_x^{LSDA}(\mathbf{r})G_{xc}^{NN}(\mathbf{g}(\mathbf{r})) \quad (2.14)$$

où $\mathbf{g}(\mathbf{r})$ est un vecteur qui contient les ingrédients des fonctionnelles. Ils ont considéré quatre approches différentes, où la flèche indique les ingrédients qui sont utilisés dans chaque méthode: $\mathbf{g}^{LSDA}(\mathbf{r}) \rightarrow [\rho(\mathbf{r}), \zeta(\mathbf{r})]$, $\mathbf{g}^{GGA}(\mathbf{r}) \rightarrow [\rho(\mathbf{r}), \zeta(\mathbf{r}), \nabla\rho(\mathbf{r})]$, $\mathbf{g}^{Meta-GGA}(\mathbf{r}) \rightarrow [\rho(\mathbf{r}), \zeta(\mathbf{r}), \nabla\rho(\mathbf{r}), \tau(\mathbf{r})]$ et $\mathbf{g}^{NRA}(\mathbf{r}) \rightarrow [\rho(\mathbf{r}), \zeta(\mathbf{r}), \nabla\rho(\mathbf{r}), \tau(\mathbf{r}), \int d\mathbf{r}_2 \rho(\mathbf{r}_2) e^{-5|\mathbf{r}_1 - \mathbf{r}_2|}]$. NRA correspond à *Near Region Approximation* et elle est une approche non locale, c'est à dire, qu'elle dépend de \mathbf{r}_1 et \mathbf{r}_2 , tandis que les autres approches utilisent les ingrédients communs des fonctionnelles locales ou semi-locales. Leur NN est un MLP composé de quatre couches et il est représenté par la formule suivante

$$G_{xc}^{NN}(\mathbf{g}(\mathbf{r})) = 1 + h_4(\dots(h_1(\mathbf{g}))) \quad (2.15)$$

$$h_i(\mathbf{x}) = \sigma^{NN}(\mathbf{W}_i \mathbf{x} + \mathbf{b}_i) \quad (2.16)$$

où \mathbf{W}_i et \mathbf{b}_i sont les poids et les biais à optimiser. σ^{NN} est la fonction d'activation et ils ont utilisé celle qui est appelée unité linéaire exponentielle (ELU)

$$\sigma^{ELU}(x) = \max(0, x) + \min(0, e^x - 1). \quad (2.17)$$

Les couches du NN ont 100 neurones. Les auteurs ont entraîné leur NN pour reproduire les énergies d'atomisation de trois molécules, ainsi que de minimiser l'erreur sur les prédictions de la densité électronique comparativement aux résultats provenant de la théorie CCSD(T). Un avantage considérable des NN est que la différentiation automatique permet d'obtenir facilement les dérivés du NN par rapport aux ingrédients. Nagai et co-auteurs ont utilisé ces dérivées pour résoudre le champ autocohérent. Ils se sont alors servis d'un algorithme Monte-Carlo pour optimiser les poids et les biais du NN.

Leurs quatre NN optimisés (NN-LDA, NN-GGA, NN-Méta-GGA, NN-NRA) montrent des résultats en thermochimie supérieurs aux fonctionnelles utilisant le même nombre d'ingrédients, bien qu'ils aient entraîné leur NN sur seulement trois molécules. Cette observation démontre la transférabilité de leur méthode pour des molécules organiques. En particulier, NN-NRA est une méthode qui se compare à une fonctionnelle hybride comme PBEh pour le calcul des énergies d'atomisation, des potentiels d'ionisation, des densités électroniques et des barrières d'activation. Le NN a donc réussi à trouver des tendances pertinentes pour reproduire l'énergie XC, à partir des ingrédients et des données théoriques ou expérimentales à reproduire.

Néanmoins, un élément important à mentionner est que les fonctionnelles NN qui sont construites à partir de données empiriques ont des erreurs moins systématiques que celles non empiriques [53]. Il est donc difficile d'analyser pourquoi de tels résultats sont obtenus. Un autre problème potentiel est que les NN ne sont pas fiables en dehors des plages de valeurs entraînées, ce qui pourrait parfois causer des erreurs importantes pour des molécules montrant des caractéristiques moins communes. Il serait intéressant de tester ces fonctionnelles

pour des molécules dont la structure électronique est difficile à prédire en DFT, comme les métaux de transition.

L'étude pionnière de Nagai et co-auteurs a été marquante pour la construction de fonctionnelles à l'aide du ML. Une étude intéressante qui a suivi est celle de chercheurs de la compagnie DeepMind publiée en 2021 [116]. Ils ont ajouté dans les données d'entraînement des systèmes avec des charges et des spins fractionnaires, dans le but d'améliorer les prédictions de la DFT concernant les systèmes fortement corrélés (comme la molécule de H_2 étirée par exemple) et/ou qui ont des délocalisations de charge (H_2^+ étirée). Leurs résultats montrent que leur NN est transférable pour ce type de système, en plus de conserver de bons résultats pour les molécules à leur position d'équilibre.

Les études précédentes montrent que beaucoup de chercheurs s'intéressent aux techniques du ML en KS-DFT, malgré que ce soit un domaine très récent. Cependant, il y a tout de même des chercheurs reconnus en DFT comme Axel D. Becke [156] qui préfèrent développer des approximations basées sur la théorie plutôt que d'ajuster des milliers de paramètres pour reproduire de grands ensembles de données. Nous allons expliquer notre philosophie concernant l'utilisation du ML en DFT à la prochaine section (section 2.5.1).

Cette section concerne majoritairement le développement de nouvelles fonctionnelles en DFT, mais le ML est aussi utilisé de plusieurs autres façons en DFT et en chimie computationnelle [150, 151]. Un exemple est la construction de champ de force en dynamique moléculaire (MD). En MD, les interactions entre les molécules sont généralement représentées par des équations simples comme le potentiel de Lennard-Jones et le potentiel de Coulomb. Ainsi, le ML peut construire des champs de force en entraînant des modèles sur des données provenant, par exemple de la DFT, afin de mieux reproduire les interactions intermoléculaires. Une fois le modèle ML entraîné, il serait alors en principe possible d'avoir une exactitude similaire à la DFT avec un temps de calcul très faible. Il est possible de simuler de gros systèmes comme des protéines indirectement par la DFT grâce au ML. Néanmoins, il est important de mentionner que le champ de force aura au minimum les mêmes limitations que la méthode théorique qui a généré les données d'entraînement. La construction de nouvelles fonctionnelles XC ne sera donc pas un domaine de recherche qui va disparaître de sitôt, le ML deviendra probablement un nouvel outil qui améliorera les approximations à l'énergie XC.

2.5.1. Notre philosophie concernant l'utilisation de l'apprentissage automatique pour approximer l'énergie d'échange-corrélation.

Comme mentionné à la section des approximations à l'énergie XC (section 1.5), il y a deux principales philosophies [16, 53, 54] concernant la construction de fonctionnelles: l'approche dite empirique et celle non empirique. La grande quantité de données disponibles, ainsi que la flexibilité du ML pour approximer des fonctions, font donc du ML une technique très

puissante pour développer des fonctionnelles empiriques. La fonctionnelle DM21 de la compagnie Deepmind [116] montre le potentiel du ML pour approximer l'énergie XC à partir de grands ensembles de données.

Notre groupe de recherche s'intéresse plutôt à développer de fonctionnelles non empiriques. Des équations mathématiques sont développées pour respecter des contraintes physiques et mathématiques, sans qu'elles soient ajustées pour reproduire de grands ensembles de données. Cependant, il est très difficile pour un théoricien de trouver des équations mathématiques flexibles qui respectent un nombre considérable de contraintes, tel que nous avons pu l'observer durant le projet de l'article 1 de cette thèse qui tente d'approximer le trou XC à partir de la densité non locale. Ainsi, le ML devient une technique puissante pour trouver des fonctions mathématiques qui respectent diverses contraintes. L'article 2 de cette thèse montre l'utilisation de NN pour construire des trous X non empiriquement à partir de contraintes mathématiques.

Notre groupe s'intéresse aussi à l'utilisation de nouveaux ingrédients pour développer des fonctionnelles, tels que la densité non locale (article 1) ou le quatrième terme du développement en série de Taylor du trou X (article 3). Il est parfois ardu de mesurer l'impact potentiel d'un ingrédient sur l'amélioration des approximations à l'énergie XC pour prédire les propriétés électroniques de nombreux systèmes chimiques. Ainsi, malgré que ce ne soit pas une approche non empirique, le ML permet de tester les effets d'un nouvel ingrédient sur les prédictions d'une fonctionnelle. L'étude de l'article 3 de cette thèse utilise le ML pour confirmer que l'utilisation du quatrième terme du développement en série de Taylor du trou d'échange diminue les erreurs concernant le calcul des énergies d'atomisation et des barrières d'activation des molécules, ce qui était difficile à accomplir sans le ML [49].

De plus, une étude dont la philosophie sur la construction de fonctionnelle XC s'approche de la nôtre est celle de Ma et co-auteurs qui a été publiée en 2022 [157]. Ils ont utilisé le ML non pas pour approximer l'énergie XC directement, mais pour construire des approximations à partir des équations mathématiques simples qui vont approximer l'énergie XC. Cette approche évite l'un des défauts du ML qu'il est presque impossible d'analyser en détail des fonctions qui ont des centaines, voir des milliers de paramètres. Les équations mathématiques simples qu'ils obtiennent sont aussi plus rapides à calculer que la plupart des techniques en ML. Il sera intéressant de voir l'impact futur que cette étude pourrait avoir sur le développement de fonctionnelles.

En résumé, notre groupe de recherche s'intéresse au ML pour construire des approximations non empiriques à l'énergie XC à partir des contraintes physiques, ainsi que d'étudier les propriétés de nouveaux ingrédients comme le quatrième terme du développement en série de Taylor du trou X. Malgré que des études, comme celle de Deepmind [116], où une fonction flexible est ajustée sur de grands ensembles de données donnent des fonctionnelles utiles pour des applications en DFT, nous nous intéressons plutôt à la compréhension de l'énergie

XC d'un point de vue fondamental. Par exemple, quel serait l'impact du quatrième terme du trou X sur les prédictions des barrières d'activation? Grâce à de nouveaux algorithmes et à des ordinateurs, dont la puissance de calcul continue d'augmenter, il sera intéressant d'étudier l'évolution future de la DFT avec le ML.

3. Objectifs de la thèse

L'objectif principal de cette thèse est de construire de nouvelles approximations de l'énergie d'échange-corrélation. Ces nouvelles approximations doivent avoir une meilleure exactitude ou être plus rapides à calculer que les méthodes courantes en DFT. De plus, un autre objectif des projets de cette thèse est d'avoir une meilleure compréhension fondamentale de la DFT. Dans la section 1.5.9 de la présente thèse, nous avons mentionné qu'il y a encore plusieurs limitations qui sont observées avec les fonctionnelles populaires en DFT. L'élimination de l'erreur d'auto-interaction et l'amélioration du calcul de la corrélation statique sont encore des sujets d'actualité, puisque ces erreurs diminuent l'exactitude des prédictions des fonctionnelles en chimie [114]. Alors, nous nous intéressons aussi à étudier des quantités comme la densité non locale ou le quatrième terme de l'expansion en série de Taylor du trou d'échange exacte ($T(\mathbf{r})$), afin d'avoir une meilleure compréhension fondamentale de ces quantités. Cette meilleure compréhension favorise la construction de nouvelles approximations plus efficaces en DFT.

Pour améliorer les approximations de l'énergie XC, nous allons nous inspirer de travaux précédents du groupe Ernzerhof, tel que les études sur la densité non locale ($\rho(\mathbf{r},u)$) [108] ou sur $T(\mathbf{r})$ [49]. Ces quantités ont peu été étudiées dans la littérature scientifique, il est donc pertinent de continuer d'investiguer leur impact potentiel sur les prédictions des fonctionnelles de l'énergie XC. Cependant, comme observés dans les études précédentes du groupe et dans le premier article de cette thèse, la densité non locale et $T(\mathbf{r})$ sont des quantités assez difficiles à utiliser pour construire des approximations. Ainsi, nous nous intéressons à de nouveaux outils qui ont récemment gagné en popularité en DFT, en particulier l'apprentissage automatique (ML). Les bibliothèques ML comme Pytorch [143] sont puissantes et faciles à utiliser. Les réseaux de neurones artificiels sont un outil particulièrement puissant pour construire des fonctions, donc nous allons nous servir de cet outil pour faciliter le développement de nos nouvelles fonctionnelles.

Références bibliographiques

- [1] W. Koch et M. C. Holthausen, *A Chemist's Guide to Density Functional Theory*. Weinheim - New York: Wiley - VCH, 2nd edition, 2001.
- [2] C. J. Cramer, *Essentials of Computational Chemistry: Theories and Models*. John Wiley & Sons, 2013.
- [3] F. Jensen, *Introduction to Computational Chemistry*. John Wiley & Sons, 2017.
- [4] A. Szabo et N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. Courier Corporation, 2012.
- [5] R. G. Parr et W. Yang, "Density Functional Theory of Atoms and Molecules," *Oxford University Press*, 1989.
- [6] M. Born et R. Oppenheimer, "Zur Quantentheorie der Molekeln," *Annalen der Physik*, vol. 389, no. 20, pp. 457–484, 1927.
- [7] M. Baer, "Introduction to the Theory of Electronic Non-Adiabatic Coupling Terms in Molecular Systems," *Physics Reports*, vol. 358, no. 2, pp. 75–142, 2002.
- [8] P. Pyykkö, "Relativistic Effects in Chemistry: More Common Than You Thought," *Annual Review of Physical Chemistry*, vol. 63, no. 1, pp. 45–64, 2012.
- [9] D. P. Tew, W. Klopper, et T. Helgaker, "Electron Correlation: The Many-Body Problem at the Heart of Chemistry," *Journal of Computational Chemistry*, vol. 28, no. 8, pp. 1307–1320, 2007.
- [10] D. R. Hartree, "The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part II. Some Results and Discussion," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 24, no. 1, p. 111–132, 1928.
- [11] V. Fock, "Näherungsmethode zur lösung des quantenmechanischen mehrkörperproblems," *Zeitschrift für Physik*, vol. 61, pp. 126–148, 1930.
- [12] J. A. Pople et R. K. Nesbet, "Self-Consistent Orbitals for Radicals," *The Journal of Chemical Physics*, vol. 22, no. 3, pp. 571–572, 12 2004.
- [13] C. C. J. Roothaan, "New Developments in Molecular Orbital Theory," *Rev. Mod. Phys.*, vol. 23, pp. 69–89, Apr 1951.
- [14] G. G. Hall, "The Molecular Orbital Theory of Chemical Valency VIII. a Method of Calculating Ionization Potentials," *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, vol. 205, no. 1083, pp. 541–552, 1951.
- [15] L. A. Curtiss, K. Raghavachari, P. C. Redfern, et J. A. Pople, "Assessment of Gaussian-2 and Density Functional Theories for the Computation of Enthalpies of Formation," *The Journal of Chemical Physics*, vol. 106, no. 3, pp. 1063–1079, 1997.
- [16] A. D. Becke, "Perspective: Fifty Years of Density-Functional Theory in Chemical Physics," *The Journal of Chemical Physics*, vol. 140, no. 18, p. 18A301, 2014.

- [17] F. Furche et J. P. Perdew, “The Performance of Semilocal and Hybrid Density Functionals in 3d Transition-Metal Chemistry,” *The Journal of Chemical Physics*, vol. 124, no. 4, p. 044103, 2006.
- [18] J. M. L. Martin, *Ab Initio Thermochemistry Beyond Chemical Accuracy for First-and Second-Row Compounds*. Dordrecht: Springer Netherlands, 1999, pp. 373–415.
- [19] R. Gebauer, M. H. Cohen, et R. Car, “A Well-Scaling Natural Orbital Theory,” *Proceedings of the National Academy of Sciences*, vol. 113, no. 46, pp. 12 913–12 918, 2016.
- [20] W. Kohn et L. J. Sham, “Self-Consistent Equations Including Exchange and Correlation Effects,” *Phys. Rev.*, vol. 140, pp. A1133–A1138, 1965.
- [21] P. Hohenberg et W. Kohn, “Inhomogeneous Electron Gas,” *Phys. Rev.*, vol. 136, pp. B864–B871, 1964.
- [22] P. Macchi, J.-M. Gillet, F. Taulelle, J. Campo, N. Claiser, et C. Lecomte, “Modelling the Experimental Electron Density: Only the Synergy of Various Approaches Can Tackle the New Challenges,” *IUCrJ*, vol. 2, no. 4, pp. 441–451, 2015.
- [23] C. F. Matta et R. J. Gillespie, “Understanding and Interpreting Molecular Electron Density Distributions,” *Journal of Chemical Education*, vol. 79, no. 9, p. 1141, 2002.
- [24] R. F. Bader, “Atoms in Molecules,” *Accounts of Chemical Research*, vol. 18, no. 1, pp. 9–15, 1985.
- [25] L. H. Thomas, “The Calculation of Atomic Fields,” *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 23, no. 5, p. 542–548, 1927.
- [26] E. Fermi, “Un metodo statistico per la determinazione di alcune priorieta dell’atome,” *Rend. Accad. Naz. Lincei*, vol. 6, no. 602-607, p. 32, 1927.
- [27] P. A. M. Dirac, “Note on Exchange Phenomena in the Thomas Atom,” *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 26, no. 3, p. 376–385, 1930.
- [28] NobelPrize.org, “Prix nobel de chimie,” 1998. [En-ligne]. Disponible: <https://www.nobelprize.org/prizes/chemistry/1998/summary/>
- [29] M. Levy, “Universal Variational Functionals of Electron Densities, First-order Density Matrices, and Natural Spin-Orbitals and Solution of the v -Representability Problem,” *Proceedings of the National Academy of Sciences*, vol. 76, no. 12, pp. 6062–6065, 1979.
- [30] E. H. Lieb, “Density Functionals For Coulomb Systems,” *International Journal of Quantum Chemistry*, vol. 24, no. 3, pp. 243–277, 1983.
- [31] D. Langreth et J. Perdew, “The Exchange-Correlation Energy of a Metallic Surface,” *Solid State Communications*, vol. 17, no. 11, pp. 1425–1429, 1975.
- [32] O. Gunnarsson et B. I. Lundqvist, “Exchange and Correlation in Atoms, Molecules, and Solids by the Spin-Density-Functional Formalism,” *Phys. Rev. B*, vol. 13, pp. 4274–4298, 1976.
- [33] A. D. Becke, “A New Mixing of Hartree–Fock and Local Density-Functional Theories,” *The Journal of Chemical Physics*, vol. 98, no. 2, pp. 1372–1377, 1993.

- [34] J. C. Slater, "A Simplification of the Hartree-Fock Method," *Phys. Rev.*, vol. 81, pp. 385–390, 1951.
- [35] Y. Wang, E. Proynov, et J. Kong, "Model DFT Exchange Holes and the Exact Exchange Hole: Similarities and Differences," *The Journal of Chemical Physics*, vol. 154, no. 2, p. 024101, 2021.
- [36] J. P. Perdew et A. Zunger, "Self-Interaction Correction to Density-Functional Approximations for Many-Electron Systems," *Phys. Rev. B*, vol. 23, pp. 5048–5079, 1981.
- [37] M. Lundberg et P. E. M. Siegbahn, "Quantifying the Effects of the Self-Interaction Error in DFT: When do the Delocalized States Appear?" *The Journal of Chemical Physics*, vol. 122, no. 22, p. 224103, 2005.
- [38] B. G. Johnson, C. A. Gonzales, P. M. Gill, et J. A. Pople, "A Density Functional Study of the Simplest Hydrogen Abstraction Reaction. Effect of Self-Interaction Correction," *Chemical Physics Letters*, vol. 221, no. 1, pp. 100–108, 1994.
- [39] H. Bahmann et M. Ernzerhof, "Generalized-Gradient Exchange-Correlation Hole Obtained from a Correlation Factor Ansatz," *The Journal of Chemical Physics*, vol. 128, no. 23, p. 234104, 2008.
- [40] J. Přečetlová, H. Bahmann, M. Kaupp, et M. Ernzerhof, "Communication: A Non-Empirical Correlation Factor Model for the Exchange-Correlation Energy," *The Journal of Chemical Physics*, vol. 141, no. 11, p. 111102, 2014.
- [41] J. P. Perdew et Y. Wang, "Pair-Distribution Function and its Coupling-Constant Average for the Spin-Polarized Electron Gas," *Phys. Rev. B*, vol. 46, pp. 12 947–12 954, 1992.
- [42] P. Gori-Giorgi et J. P. Perdew, "Pair Distribution Function of the Spin-Polarized Electron Gas: A First-Principles Analytic Model for All Uniform Densities," *Phys. Rev. B*, vol. 66, p. 165118, 2002.
- [43] J. Pavlíková Přečetlová, H. Bahmann, M. Kaupp, et M. Ernzerhof, "Design of Exchange-Correlation Functionals Through the Correlation Factor Approach," *The Journal of Chemical Physics*, vol. 143, no. 14, p. 144102, 2015.
- [44] R. Wang, Y. Zhou, et M. Ernzerhof, "Construction of Self-Interaction-Corrected Exchange-Correlation Functionals Within the Correlation Factor Approach," *The Journal of Chemical Physics*, vol. 151, no. 19, p. 194102, 2019.
- [45] R. Wang, Y. Zhou, et M. Ernzerhof, "The Correlation Factor Model for the Exchange-Correlation Energy and its Application to Transition Metal Compounds," *The Journal of Chemical Physics*, vol. 150, no. 8, p. 084107, 2019.
- [46] P.-O. Roy, E. Cuierrier, et M. Ernzerhof, "The Correlation Factor Approach: Combining Density Functional and Wave Function Theory," *The Journal of Chemical Physics*, vol. 152, no. 21, p. 211101, 2020.

- [47] A. D. Becke, "Hartree–Fock Exchange Energy of an Inhomogeneous Electron Gas," *International Journal of Quantum Chemistry*, vol. 23, no. 6, pp. 1915–1922, 1983.
- [48] A. D. Becke et M. R. Roussel, "Exchange Holes in Inhomogeneous Systems: A Coordinate-Space Model," *Phys. Rev. A*, vol. 39, pp. 3761–3767, 1989.
- [49] R. Wang, Y. Zhou, et M. Ernzerhof, "Fourth-Order Series Expansion of the Exchange Hole," *Phys. Rev. A*, vol. 96, p. 022502, 2017.
- [50] A. D. Becke, "A real-space model of nondynamical correlation," *The Journal of Chemical Physics*, vol. 119, no. 6, pp. 2972–2977, 2003.
- [51] E. J. Baerends, "Exact Exchange-Correlation Treatment of Dissociated H_2 in Density Functional Theory," *Phys. Rev. Lett.*, vol. 87, p. 133004, 2001.
- [52] A. D. Becke, "Real-Space Post-Hartree–Fock Correlation Models," *The Journal of Chemical Physics*, vol. 122, no. 6, p. 064101, 2005.
- [53] D. Rappoport, N. R. Crawford, F. Furche, K. Burke, et C. Wiley, "Which Functional Should I Choose?" *Computational Inorganic and Bioinorganic Chemistry*, Wiley-Blackwell, 2009.
- [54] K. Burke, "Perspective on Density Functional Theory," *The Journal of Chemical Physics*, vol. 136, no. 15, p. 150901, 2012.
- [55] M. G. Medvedev, I. S. Bushmarinov, J. Sun, J. P. Perdew, et K. A. Lyssenko, "Density Functional Theory is Straying from the Path Toward the Exact Functional," *Science*, vol. 355, no. 6320, pp. 49–52, 2017.
- [56] J. P. Perdew et K. Schmidt, "Jacob's Ladder of Density Functional Approximations for the Exchange–Correlation Energy," *AIP Conference Proceedings*, vol. 577, no. 1, pp. 1–20, 2001.
- [57] J. P. Perdew, K. Burke, et M. Ernzerhof, "Generalized Gradient Approximation Made Simple," *Phys. Rev. Lett.*, vol. 77, pp. 3865–3868, 1996.
- [58] J. Sun, A. Ruzsinszky, et J. P. Perdew, "Strongly Constrained and Appropriately Normed Semilocal Density Functional," *Phys. Rev. Lett.*, vol. 115, p. 036402, 2015.
- [59] A. D. Becke, "Density-Functional Thermochemistry. III. The Role of Exact Exchange," *The Journal of Chemical Physics*, vol. 98, no. 7, pp. 5648–5652, 1993.
- [60] C. Lee, W. Yang, et R. G. Parr, "Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density," *Phys. Rev. B*, vol. 37, pp. 785–789, 1988.
- [61] J. P. Perdew, M. Ernzerhof, et K. Burke, "Rationale for Mixing Exact Exchange with Density Functional Approximations," *The Journal of Chemical Physics*, vol. 105, no. 22, pp. 9982–9985, 1996.
- [62] C. Adamo et V. Barone, "Toward Reliable Density Functional Methods Without Adjustable Parameters: The PBE0 Model," *The Journal of Chemical Physics*, vol. 110, no. 13, pp. 6158–6170, 1999.

- [63] S. Grimme et F. Neese, “Double-Hybrid Density Functional Theory for Excited Electronic States of Molecules,” *The Journal of Chemical Physics*, vol. 127, no. 15, p. 154116, 2007.
- [64] P.-O. Roy, E. Cuierrier, et M. Ernzerhof, “The Correlation Factor Approach: Combining Density Functional and Wave Function Theory,” *The Journal of Chemical Physics*, vol. 152, no. 21, p. 211101, 2020.
- [65] M. Ernzerhof et J. P. Perdew, “Generalized Gradient Approximation to the Angle- and System-Averaged Exchange Hole,” *The Journal of Chemical Physics*, vol. 109, no. 9, pp. 3313–3320, 1998.
- [66] D. M. Ceperley et B. J. Alder, “Ground State of the Electron Gas by a Stochastic Method,” *Phys. Rev. Lett.*, vol. 45, pp. 566–569, 1980.
- [67] J. P. Perdew et Y. Wang, “Accurate and Simple Analytic Representation of the Electron-Gas Correlation Energy,” *Phys. Rev. B*, vol. 45, pp. 13 244–13 249, 1992.
- [68] S. H. Vosko, L. Wilk, et M. Nusair, “Accurate Spin-Dependent Electron Liquid Correlation Energies for Local Spin Density Calculations: A Critical Analysis,” *Canadian Journal of Physics*, vol. 58, no. 8, pp. 1200–1211, 1980.
- [69] M. Ernzerhof, J. P. Perdew, et K. Burke, *Density Functionals: Where do They Come From, Why do They Work?* Berlin, Heidelberg: Springer Berlin Heidelberg, 1996, pp. 1–30.
- [70] A. V. Arbuznikov et M. Kaupp, “Local Hybrid Exchange-Correlation Functionals Based on the Dimensionless Density Gradient,” *Chemical Physics Letters*, vol. 440, no. 1, pp. 160–168, 2007.
- [71] E. H. Lieb et S. Oxford, “Improved Lower Bound on the Indirect Coulomb Energy,” *International Journal of Quantum Chemistry*, vol. 19, no. 3, pp. 427–439, 1981.
- [72] A. D. Becke, “Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior,” *Phys. Rev. A*, vol. 38, pp. 3098–3100, 1988.
- [73] P. Elliott et K. Burke, “Non-Empirical Derivation of the Parameter in the B88 Exchange Functional,” *Canadian Journal of Chemistry*, vol. 87, no. 10, pp. 1485–1491, 2009.
- [74] Y. Zhang et W. Yang, “Comment on “Generalized Gradient Approximation Made Simple”,” *Phys. Rev. Lett.*, vol. 80, pp. 890–890, 1998.
- [75] S. Dick et M. Fernandez-Serra, “Highly Accurate and Constrained Density Functional Obtained with Differentiable Programming,” *Physical Review B*, vol. 104, no. 16, p. L161109, 2021.
- [76] J. P. Perdew, S. Kurth, A. Zupan, et P. Blaha, “Accurate Density Functional with Correct Formal Properties: A Step Beyond the Generalized Gradient Approximation,” *Phys. Rev. Lett.*, vol. 82, pp. 2544–2547, 1999.

- [77] J. Tao, J. P. Perdew, V. N. Staroverov, et G. E. Scuseria, “Climbing the Density Functional Ladder: Nonempirical Meta-Generalized Gradient Approximation Designed for Molecules and Solids,” *Phys. Rev. Lett.*, vol. 91, p. 146401, 2003.
- [78] Y. Zhao et D. G. Truhlar, “Applications and Validations of the Minnesota Density Functionals,” *Chemical Physics Letters*, vol. 502, no. 1, pp. 1–13, 2011.
- [79] H. S. Yu, X. He, et D. G. Truhlar, “MN15-L: A New Local Exchange-Correlation Functional for Kohn-Sham Density Functional Theory with Broad Accuracy for Atoms, Molecules, and Solids,” *Journal of Chemical Theory and Computation*, vol. 12, no. 3, pp. 1280–1293, 2016.
- [80] M. Kaupp, H. Bahmann, et A. V. Arbuznikov, “Local Hybrid Functionals: An Assessment for Thermochemical Kinetics,” *The Journal of Chemical Physics*, vol. 127, no. 19, p. 194102, 2007.
- [81] T. M. Maier, A. V. Arbuznikov, et M. Kaupp, “Local Hybrid Functionals: Theory, Implementation, and Performance of an Emerging New Tool in Quantum Chemistry and Beyond,” *WIREs Computational Molecular Science*, vol. 9, no. 1, p. e1378, 2019.
- [82] J. Jaramillo, G. E. Scuseria, et M. Ernzerhof, “Local Hybrid Functionals,” *The Journal of Chemical Physics*, vol. 118, no. 3, pp. 1068–1073, 2003.
- [83] H. Bahmann, A. Rodenberg, A. V. Arbuznikov, et M. Kaupp, “A Thermochemically Competitive Local Hybrid Functional Without Gradient Corrections,” *The Journal of Chemical Physics*, vol. 126, no. 1, p. 011103, 2007.
- [84] T. M. Henderson, B. G. Janesko, et G. E. Scuseria, “Range Separation and Local Hybridization in Density Functional Theory,” *The Journal of Physical Chemistry A*, vol. 112, no. 49, pp. 12 530–12 542, 2008.
- [85] H. Bahmann, “From Global to Local – Hybrid Density Functionals for Weak and Strong Correlation,” dans *Chemical Modelling: Volume 16*. The Royal Society of Chemistry, 2021.
- [86] B. G. Janesko, “Replacing Hybrid Density Functional Theory: Motivation and Recent Advances,” *Chem. Soc. Rev.*, vol. 50, pp. 8470–8495, 2021.
- [87] H. Stoll et A. Savin, *Density Functionals for Correlation Energies of Atoms and Molecules*. Boston, MA: Springer US, 1985, pp. 177–207.
- [88] A. Savin et H.-J. Flad, “Density Functionals for the Yukawa Electron-Electron Interaction,” *International Journal of Quantum Chemistry*, vol. 56, no. 4, pp. 327–332, 1995.
- [89] A. Savin, “On Degeneracy, Near-Degeneracy and Density Functional Theory,” dans *Recent Developments and Applications of Modern Density Functional Theory*, ser. Theoretical and Computational Chemistry, J. Seminario, Ed. Elsevier, 1996, vol. 4, pp. 327–357.

- [90] J. Heyd, G. E. Scuseria, et M. Ernzerhof, “Hybrid Functionals Based on a Screened Coulomb Potential,” *The Journal of Chemical Physics*, vol. 118, no. 18, pp. 8207–8215, 2003.
- [91] J. Toulouse, F. Colonna, et A. Savin, “Long-Range–Short-Range Separation of the Electron-Electron Interaction in Density-Functional Theory,” *Phys. Rev. A*, vol. 70, p. 062505, 2004.
- [92] T. Yanai, D. P. Tew, et N. C. Handy, “A New Hybrid Exchange–Correlation Functional Using the Coulomb-Attenuating Method (CAM-B3LYP),” *Chemical Physics Letters*, vol. 393, no. 1, pp. 51–57, 2004.
- [93] A. V. Krukau, G. E. Scuseria, J. P. Perdew, et A. Savin, “Hybrid Functionals with Local Range Separation,” *The Journal of Chemical Physics*, vol. 129, no. 12, p. 124103, 2008.
- [94] R. Peverati et D. G. Truhlar, “Quest for a Universal Density Functional: The Accuracy of Density Functionals Across a Broad Spectrum of Databases in Chemistry and Physics,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 372, no. 2011, p. 20120476, 2014.
- [95] E. Brémond, A. J. Pérez-Jiménez, J. C. Sancho-García, et C. Adamo, “Range-Separated Hybrid Density Functionals Made Simple,” *The Journal of Chemical Physics*, vol. 150, no. 20, p. 201102, 2019.
- [96] A. Savin, “Models and Corrections: Range Separation for Electronic Interaction-Lessons from Density Functional Theory,” *The Journal of Chemical Physics*, vol. 153, no. 16, p. 160901, 2020.
- [97] S. Klawohn et H. Bahmann, “Self-Consistent Implementation of Hybrid Functionals with Local Range Separation,” *Journal of Chemical Theory and Computation*, vol. 16, no. 2, pp. 953–963, 2020.
- [98] M. Brütting, H. Bahmann, et S. Kümmel, “Hybrid Functionals with Local Range Separation: Accurate Atomization Energies and Reaction Barrier Heights,” *The Journal of Chemical Physics*, vol. 156, no. 10, p. 104109, 2022.
- [99] G. Santra, R. Calinsky, et J. M. L. Martin, “Benefits of Range-Separated Hybrid and Double-Hybrid Functionals for a Large and Diverse Data Set of Reaction Energies and Barrier Heights,” *The Journal of Physical Chemistry A*, vol. 126, no. 32, pp. 5492–5505, 2022.
- [100] O. A. Vydrov et G. E. Scuseria, “Assessment of a Long-Range Corrected Hybrid Functional,” *The Journal of Chemical Physics*, vol. 125, no. 23, p. 234109, 2006.
- [101] J. E. Moussa, P. A. Schultz, et J. R. Chelikowsky, “Analysis of the Heyd-Scuseria-Ernzerhof Density Functional Parameter Space,” *The Journal of Chemical Physics*, vol. 136, no. 20, p. 204117, 2012.

- [102] T. Aschebrock et S. Kümmel, “Exploring Local Range Separation: The Role of Spin Scaling and One-Electron Self-Interaction,” *The Journal of Chemical Physics*, vol. 151, no. 15, p. 154108, 2019.
- [103] S. Klawohn et H. Bahmann, “Self-Consistent Implementation of Hybrid Functionals with Local Range Separation,” *Journal of Chemical Theory and Computation*, vol. 16, no. 2, pp. 953–963, 2020.
- [104] T. M. Maier, Y. Iwabata, et H. Nakai, “Assessing Locally Range-Separated Hybrid Functionals from a Gradient Expansion of the Exchange Energy Density,” *The Journal of Chemical Physics*, vol. 154, no. 21, p. 214101, 2021.
- [105] O. Gunnarsson, M. Jonson, et B. Lundqvist, “Exchange and Correlation in Inhomogeneous Electron Systems,” *Solid State Communications*, vol. 24, no. 11, pp. 765–768, 1977.
- [106] J. A. Alonso et L. A. Girifalco, “Nonlocal Approximation to the Exchange Potential and Kinetic Energy of an Inhomogeneous Electron Gas,” *Phys. Rev. B*, vol. 17, pp. 3735–3743, 1978.
- [107] O. Gunnarsson, M. Jonson, et B. I. Lundqvist, “Descriptions of Exchange and Correlation Effects in Inhomogeneous Electron Systems,” *Phys. Rev. B*, vol. 20, pp. 3136–3164, 1979.
- [108] H. Antaya, Y. Zhou, et M. Ernzerhof, “Approximating the Exchange Energy Through the Nonempirical Exchange-Factor Approach,” *Phys. Rev. A*, vol. 90, p. 032513, 2014.
- [109] R. Cuevas-Saavedra, D. Chakraborty, S. Rabi, C. Cárdenas, et P. W. Ayers, “Symmetric Nonlocal Weighted Density Approximations from the Exchange-Correlation Hole of the Uniform Electron Gas,” *Journal of Chemical Theory and Computation*, vol. 8, no. 11, pp. 4081–4093, 2012.
- [110] R. Cuevas-Saavedra, D. Chakraborty, et P. W. Ayers, “Symmetric Two-Point Weighted Density Approximation for Exchange Energies,” *Phys. Rev. A*, vol. 85, p. 042519, 2012.
- [111] P. García-González, J. E. Alvarillos, E. Chacón, et P. Tarazona, “Image Potential and the Exchange-Correlation Weighted Density Approximation Functional,” *Phys. Rev. B*, vol. 62, pp. 16 063–16 068, 2000.
- [112] H. Bahmann, Y. Zhou, et M. Ernzerhof, “The Shell Model for the Exchange-Correlation Hole in the Strong-Correlation Limit,” *The Journal of Chemical Physics*, vol. 145, no. 12, p. 124104, 2016.
- [113] P. Gori-Giorgi, J. G. Ángyán, et A. Savin, “Charge Density Reconstitution from Approximate Exchange-Correlation Holes,” *Canadian Journal of Chemistry*, vol. 87, no. 10, pp. 1444–1450, 2009.
- [114] A. J. Cohen, P. Mori-Sánchez, et W. Yang, “Insights into Current Limitations of Density Functional Theory,” *Science*, vol. 321, no. 5890, pp. 792–794, 2008.

- [115] J. P. Perdew, R. G. Parr, M. Levy, et J. L. Balduz, “Density-Functional Theory for Fractional Particle Number: Derivative Discontinuities of the Energy,” *Phys. Rev. Lett.*, vol. 49, pp. 1691–1694, 1982.
- [116] J. Kirkpatrick, B. McMorrow, D. H. P. Turban, A. L. Gaunt, J. S. Spencer, A. G. D. G. Matthews, A. Obika, L. Thiry, M. Fortunato, D. Pfau, L. R. Castellanos, S. Petersen, A. W. R. Nelson, P. Kohli, P. Mori-Sánchez, D. Hassabis, et A. J. Cohen, “Pushing the Frontiers of Density Functionals by Solving the Fractional Electron Problem,” *Science*, vol. 374, no. 6573, pp. 1385–1389, 2021.
- [117] Q. Wu et W. Yang, “Empirical Correction to Density Functional Theory for van der Waals Interactions,” *The Journal of Chemical Physics*, vol. 116, no. 2, pp. 515–524, 2002.
- [118] S. Grimme, J. Antony, S. Ehrlich, et H. Krieg, “A Consistent and Accurate ab initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-PU,” *The Journal of Chemical Physics*, vol. 132, no. 15, p. 154104, 2010.
- [119] K. Berland, V. R. Cooper, K. Lee, E. Schröder, T. Thonhauser, P. Hyldgaard, et B. I. Lundqvist, “van der Waals Forces in Density Functional Theory: A Review of the vdW-DF Method,” *Reports on Progress in Physics*, vol. 78, no. 6, p. 066501, 2015.
- [120] O. A. Vydrov et T. Van Voorhis, “Nonlocal van der Waals Density Functional: The Simpler the Better,” *The Journal of Chemical Physics*, vol. 133, no. 24, p. 244103, 2010.
- [121] W. Hujo et S. Grimme, “Performance of the van der Waals Density Functional VV10 and (Hybrid) GGA Variants for Thermochemistry and Noncovalent Interactions,” *Journal of Chemical Theory and Computation*, vol. 7, no. 12, pp. 3866–3871, 2011.
- [122] A. L. Magalhães, “Gaussian-Type Orbitals Versus Slater-Type Orbitals: A Comparison,” *Journal of Chemical Education*, vol. 91, no. 12, pp. 2124–2127, 2014.
- [123] M. I. Jordan et T. M. Mitchell, “Machine Learning: Trends, Perspectives, and Prospects,” *Science*, vol. 349, no. 6245, pp. 255–260, 2015.
- [124] I. H. Sarker, “Machine Learning: Algorithms, Real-World Applications and Research Directions,” *SN Computer Science*, vol. 2, no. 3, p. 160, 2021.
- [125] Y. LeCun, Y. Bengio, et G. Hinton, “Deep Learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [126] T. Susnjak, “ChatGPT: The End of Online Exam Integrity?” 2022.
- [127] C. A. Gao, F. M. Howard, N. S. Markov, E. C. Dyer, S. Ramesh, Y. Luo, et A. T. Pearson, “Comparing Scientific Abstracts Generated by CHATGPT to Original Abstracts Using an Artificial Intelligence Output Detector, Plagiarism detector, and Blinded Human Reviewers,” *bioRxiv*, 2022.
- [128] U. Hasson, S. A. Nastase, et A. Goldstein, “Direct-Fit to Nature: An Evolutionary Perspective on Biological (and Artificial) Neural Networks,” *bioRxiv*, 2019.

- [129] E. Stevens, L. Antiga, et T. Viehmann, *Deep Learning With PyTorch*. Manning Publications, 2020.
- [130] I. J. Goodfellow, Y. Bengio, et A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016, <http://www.deeplearningbook.org>.
- [131] S. Manzhos et T. Carrington, “Neural Network Potential Energy Surfaces for Small Molecules and Reactions,” *Chemical Reviews*, vol. 121, no. 16, pp. 10 187–10 217, 2021.
- [132] T. Bikku, R. A. Fritz, Y. J. Colón, et F. Herrera, “Machine Learning Identification of Organic Compounds Using Visible Light,” *The Journal of Physical Chemistry A*, vol. 127, no. 10, pp. 2407–2414, 2023.
- [133] E. G. Dada, J. S. Bassi, H. Chiroma, S. M. Abdulhamid, A. O. Adetunmbi, et O. E. Ajibuwa, “Machine Learning for Email Spam Filtering: Review, Approaches and Open Research Problems,” *Heliyon*, vol. 5, no. 6, p. e01802, 2019.
- [134] W. A. Awad et S. ELseuofi, “Machine Learning Methods for Spam e-mail Classification,” *Int. J. Comput. Sci. Inf. Technol*, vol. 3, no. 1, pp. 173–184, 2011.
- [135] C. S. Adorf, T. C. Moore, Y. J. U. Melle, et S. C. Glotzer, “Analysis of Self-Assembly Pathways with Unsupervised Machine Learning Algorithms,” *The Journal of Physical Chemistry B*, vol. 124, no. 1, pp. 69–78, 2020.
- [136] C. Duan, F. Liu, A. Nandy, et H. J. Kulik, “Semi-supervised Machine Learning Enables the Robust Detection of Multireference Character at Low Cost,” *The Journal of Physical Chemistry Letters*, vol. 11, no. 16, pp. 6640–6648, 2020.
- [137] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, et M. Riedmiller, “Playing Atari with Deep Reinforcement Learning,” 2013.
- [138] K. Ahuja, W. H. Green, et Y.-P. Li, “Learning to Optimize Molecular Geometries Using Reinforcement Learning,” *Journal of Chemical Theory and Computation*, vol. 17, no. 2, pp. 818–825, 2021.
- [139] K. Hornik, M. Stinchcombe, et H. White, “Multilayer Feedforward Networks are Universal Approximators,” *Neural Networks*, vol. 2, no. 5, pp. 359–366, 1989.
- [140] A. LeNail, “NN-SVG: Publication-Ready Neural Network Architecture Schematics,” *Journal of Open Source Software*, vol. 4, no. 33, p. 747, 2019.
- [141] H. J. Kelley, “Gradient Theory of Optimal Flight Paths,” *Ars Journal*, vol. 30, no. 10, pp. 947–954, 1960.
- [142] D. E. Rumelhart, G. E. Hinton, et R. J. Williams, “Learning Representations by Back-Propagating Errors,” *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.
- [143] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, et S. Chintala, “PyTorch: An Imperative Style, High-Performance Deep Learning Library,” dans *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc., 2019, pp. 8024–8035.

- [144] D. P. Kingma et J. Ba, “Adam: A Method for Stochastic Optimization,” 2014.
- [145] C. G. Broyden, “The Convergence of a Class of Double-rank Minimization Algorithms 1. General Considerations,” *IMA Journal of Applied Mathematics*, vol. 6, no. 1, pp. 76–90, 1970.
- [146] R. Fletcher, “A New Approach to Variable Metric Algorithms,” *The Computer Journal*, vol. 13, no. 3, pp. 317–322, 1970.
- [147] D. Goldfarb, “A Family of Variable-Metric Methods Derived by Variational Means,” *Mathematics of Computation*, vol. 24, no. 109, pp. 23–26, 1970.
- [148] D. F. Shanno, “Conditioning of Quasi-Newton Methods for Function Minimization,” *Mathematics of Computation*, vol. 24, no. 111, pp. 647–656, 1970.
- [149] B. Kalita, L. Li, R. J. McCarty, et K. Burke, “Learning to Approximate Density Functionals,” *Accounts of Chemical Research*, vol. 54, no. 4, pp. 818–826, 2021.
- [150] S. Manzhos, “Machine Learning for the Solution of the Schrödinger Equation,” *Machine Learning: Science and Technology*, vol. 1, no. 1, p. 013002, 2020.
- [151] J. A. Keith, V. Vassilev-Galindo, B. Cheng, S. Chmiela, M. Gastegger, K.-R. Müller, et A. Tkatchenko, “Combining Machine Learning and Computational Chemistry for Predictive Insights Into Chemical Systems,” *arXiv:2102.06321 [physics]*, 2021.
- [152] J. Westermayr, M. Gastegger, K. T. Schütt, et R. J. Maurer, “Perspective on Integrating Machine Learning into Computational Chemistry and Materials Science,” *The Journal of Chemical Physics*, vol. 154, no. 23, p. 230903, 2021.
- [153] L. Fiedler, K. Shah, M. Bussmann, et A. Cangi, “Deep Dive into Machine Learning Density Functional Theory for Materials Science and Chemistry,” *Physical Review Materials*, vol. 6, no. 4, p. 040301, 2022.
- [154] J. C. Snyder, M. Rupp, K. Hansen, K.-R. Müller, et K. Burke, “Finding Density Functionals with Machine Learning,” *Physical Review Letters*, vol. 108, no. 25, p. 253002, 2012.
- [155] R. Nagai, R. Akashi, et O. Sugino, “Completing Density Functional Theory by Machine Learning Hidden Messages from Molecules,” *npj Computational Materials*, vol. 6, no. 1, p. 43, 2020.
- [156] A. D. Becke, “Density-Functional Theory vs Density-Functional Fits,” *The Journal of Chemical Physics*, vol. 156, no. 21, p. 214101, 2022.
- [157] H. Ma, A. Narayanaswamy, P. Riley, et L. Li, “Evolving Symbolic Density Functionals,” *Science Advances*, vol. 8, no. 36, p. eabq0279, 2022.

Premier article.

The factorization ansatz for non-local approximations to the exchange-correlation hole

par

Etienne Cuierrier¹, Pierre-Olivier Roy¹, Matthias Ernzerhof¹

(¹) Département de Chimie, Université de Montréal, C.P. 6128 Succursale A, Montréal, Québec H3C 3J7, Canada

Cet article a été publié dans *The Journal of Chemical Physics*, 156, 184110 (2022); <https://doi.org/10.1063/5.0077287>.

Les principales contributions de Étienne Cuierrier à cet article sont présentées. J'ai fait la construction, l'implémentation et les calculs associés aux modèles. J'ai aussi contribué à l'écriture de l'article et à l'analyse des résultats. Le co-auteur Pierre-Olivier Roy a écrit le code pour le calcul des densités d'énergie multiconfigurationnelle et d'échange exact. Yongxi Zhou (il est dans les remerciements de l'article) a écrit le code pour le calcul de la densité électronique non locale.

Cet article concerne le développement de nouvelles approximations non locales de l'énergie d'échange-corrélation en théorie de la fonctionnelle de la densité. Ce projet est une suite de travaux précédents du groupe: le facteur de corrélation et le facteur d'échange. Le facteur de corrélation a la limitation de ne pas reproduire correctement la limite des systèmes à un électron (erreur d'auto-interaction à un électron), tels que l'atome d'hydrogène. Cette limitation cause une détérioration des prédictions pour divers systèmes chimiques, comme les états de transition. Ainsi, puisque les modèles non locaux comme le facteur d'échange peuvent éliminer l'erreur d'auto-interaction à un électron, cet article s'intéresse à combiner le facteur de corrélation avec le facteur d'échange, afin de construire un facteur d'échange-corrélation.

Dans notre étude, nous avons construit plusieurs facteurs d'échange-corrélation qui sont exacts pour les systèmes à un électron. Néanmoins, les erreurs absolues moyennes pour le calcul d'énergie d'atomisation d'un ensemble de données de molécules, avec notre meilleur modèle non empirique, ne sont pas significativement plus faibles que celles pour une fonctionnelle semi-locale comme PBE. Nous avons étudié graphiquement nos facteurs d'échange-corrélation pour la molécule de N_2 et nous montrons que la densité électronique non locale a un comportement fortement oscillatoire. Ces nombreuses oscillations de la densité complexifient la construction du facteur d'échange-corrélation. Ainsi, pour de travaux futurs, il est nécessaire de trouver un modèle assez flexible pour s'adapter aux nombreuses oscillations de la densité non locale.

ABSTRACT.

Among the various types of approximations to the exchange-correlation energy (E_{XC}), the completely non-local approach is one of the lesser explored approximation schemes. It has not yet reached the predictive power of the widely used generalized gradient approximations, meta-generalized gradient approximations, hybrids etc. In non-local functionals, the electron density at every point in space is employed to express the exchange-correlation energy per particle $\epsilon_{XC}(\mathbf{r})$ at a given position \mathbf{r} . Here we use the non-local, spherical-averaged density $\rho(\mathbf{r},u) = \int \frac{d\Omega_{\mathbf{u}}}{4\pi} \rho(\mathbf{r} + \mathbf{u})$ as a starting point to construct approximate exchange-correlation holes through the factorization ansatz $\rho_{XC}(\mathbf{r},u) = f(\mathbf{r},u)\rho(\mathbf{r},u)$. We present upper and lower bounds to the exchange energy per particle $\epsilon_X(\mathbf{r})$ in terms of $\rho(\mathbf{r},u)$. The factor $f(\mathbf{r},u)$ is then designed to satisfy various conditions which represent important exchange and correlation effects. We assess the resulting approximations and find that the complex, oscillatory structure of $\rho(\mathbf{r},u)$ makes the construction of a corresponding $f(\mathbf{r},u)$ very challenging. This conclusion, identifying the main issue of the non-local approximation, is supported by a detailed analysis of the resulting exchange-correlation holes.

Keywords: Weighted density approximations, Exchange correlation functionals, Multi-configurational self-consistent field, Generalized gradient approximations, Density functional theory

1. Introduction

The most widespread approach of density functional theory (DFT) [1–4] is the Kohn-Sham (KS) method [1], where the exchange and correlation (XC) energy E_{XC} is the crucial quantity, its accurate representation determining the reliability of the KS calculations. The development of further improved representations of E_{XC} is an ongoing pursuit and here, we add to these developments by considering completely non-local approximations where the entire electron density $\rho(\mathbf{r})$ at any value of the spatial coordinate \mathbf{r} is employed to model the exchange-correlation energy per particle $\epsilon_{\text{XC}}(\mathbf{r})$. The non-local models emerged in the 1970s, the first being the weighted density approximation (WDA) [5–7]. This approximation has been applied to both solids, e.g., [8–11] and molecules, examples include [12–16]. Non-local models such as WDA can easily be made one-electron self-interaction error (SIE) free, thus exact for systems such as H_2^+ . The SIE is often associated with an underestimation of reaction barriers, and its elimination is of interest [17]. Recently, we developed various approximations that are based on the correlation factor (CF) ansatz [18–22] and one remaining problem with these CF functionals is the SIE. Part of the motivation of the present work is to extend the CF method to a completely non-local ansatz that allows the elimination of the SIE. Similarly, the completely non-local ansatz also enables us to consider the constraint that the X hole ($\rho_{\text{X}}(\mathbf{r}, \mathbf{r}')$) and the XC hole ($\rho_{\text{XC}}(\mathbf{r}, \mathbf{r}')$) of KS theory is bound from below by $-\rho(\mathbf{r}')$. While we found it to be very difficult to explicitly build this condition into approximations, we observe that it is satisfied by the proposed holes for the systems studied. Essentially, we attempt to build more non-locality into the X hole model which allows us to consider additional constraints. This additional non-locality has then to be addressed through the design of appropriate correlation functionals.

Before turning to the subject of the article, we discuss the early version of the WDA. In this approach E_{XC} is customarily written as

$$E_{\text{XC}} = \frac{1}{2} \int \int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')(\bar{g}(\mathbf{r}, \mathbf{r}') - 1)}{|\mathbf{r} - \mathbf{r}'|} d^3r' d^3r = \frac{1}{2} \int \int \rho(\mathbf{r}) \frac{\rho_{xc}(\mathbf{r}, \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d^3r' d^3r, \quad (1.1)$$

where $\bar{g}(\mathbf{r}, \mathbf{r}')$ is the electronic pair-correlation function integrated over the Coulomb interaction strength λ ($\bar{g}(\mathbf{r}, \mathbf{r}') = \int_0^1 d\lambda g_\lambda(\mathbf{r}, \mathbf{r}')$) and $\rho_{\text{XC}}(\mathbf{r}, u)$ is the exchange-correlation hole. $\bar{g}(\mathbf{r}, \mathbf{r}')$ satisfies various constraints such as the normalization condition,

$$\int (\bar{g}(\mathbf{r}, \mathbf{r}') - 1)\rho(\mathbf{r}') d^3r' = -1 \quad (1.2)$$

and symmetry condition

$$\bar{g}(\mathbf{r}, \mathbf{r}') = \bar{g}(\mathbf{r}', \mathbf{r}). \quad (1.3)$$

In most WDA approximation, the pair-correlation function of the homogeneous electron gas (HEG) is employed and a weighted density ($\bar{\rho}$) is determined to satisfy the normalization

condition,

$$\bar{g}(\mathbf{r}, \mathbf{r}') \approx \bar{g}^{HEG}(\mathbf{r}, \mathbf{r}'; \bar{\rho}(\mathbf{r})). \quad (1.4)$$

Recently, the symmetry condition attracted attention [8, 10, 14, 15, 23], its satisfaction is quite involved and requires the evaluation of six dimensional integrals.

In this work, we revisit the non-local approximation employing advances that have become available recently [18–22] and we systematically investigate if further improvements can be made. In detail, drawing on our earlier work [16], we employ the spherically-averaged electron density,

$$\rho(\mathbf{r}, u) = \int \frac{d\Omega_{\mathbf{u}}}{4\pi} \rho(\mathbf{r} + \mathbf{u}), \quad (1.5)$$

where \mathbf{r} is a reference position and u the electron-electron separation. $\rho(\mathbf{r}, u)$ represents the angle average of the electronic density over all the orientation of u . This has the advantage that at a given position \mathbf{r} two dimensions are being averaged over and the density merely depends on u . Furthermore, similar to the correlation factor model [18–22], we employ an exchange factor $f_X(\mathbf{r}, u)$ [16] that turns the spherically-averaged electron density into an exchange hole $\rho_X(\mathbf{r}, u)$; for a spin σ channel, we use the ansatz,

$$\rho_{X,\sigma}(\mathbf{r}, u) = f_{X,\sigma}(\mathbf{r}, u) \rho_{\sigma}(\mathbf{r}, u). \quad (1.6)$$

The subsequent application of a correlation factor [18–22] transforms the spin-averaged X-hole into an XC-hole,

$$\rho_{XC}(\mathbf{r}, u) = f_C(\mathbf{r}, u) \rho_X(\mathbf{r}, u). \quad (1.7)$$

The XC energy is then obtained from

$$E_{XC}[\rho] = \frac{1}{2} \int \int \frac{\rho(\mathbf{r}) \rho_{XC}(\mathbf{r}, u)}{u} du^3 d^3r. \quad (1.8)$$

While working with the spherically averaged hole has the advantage of eliminating the two angular variables (compare Eq. 1.1), the symmetry of the pair density with respect to permutation of the electronic coordinates cannot be preserved [24].

In recent years, $\rho(\mathbf{r}, u)$ has also been employed in various approximations to the strictly-correlated limit of DFT [25]. In the strictly correlated limit, the coupling constant of the adiabatic connection is taken to infinity and the kinetic energy of the electrons becomes negligible relative to the potential energy. In general, E_{XC} of the strictly correlated limit is difficult to calculate and various approximations have been developed to approximate this quantity in terms of $\rho(\mathbf{r}, u)$. In particular, the local radius model [26] and generalizations thereof [27] have been used that are constructed from $\rho(\mathbf{r}, u)$. Similarly, $\rho(\mathbf{r}, u)$ has also been employed to obtain approximations to E_{XC} that are not related to WDA [28]. Generally, the non-local electron density is the fundamental variable of DFT and it is of theoretical and practical interest to explore the limits of non-local approximations.

One of the new elements that we add to the WDA approach here is to impose the condition that the model X-hole, obtained from the exchange-factor approach, reproduces the exact exchange energy per particle $\epsilon_X^{exact}(\mathbf{r})$. Recently, we argued [18, 19] that in an approach that is based on exact exchange, the corresponding exact X hole is a less than ideal starting point for the subsequent inclusion of correlation. The reason being that the exact X hole shows system dependent oscillations that are extremely difficult to deal with in approximations for correlation. A simpler model X-hole, which smooths out the oscillations, is required for the modeling of correlation effects. Here we start from the non-local density to model ρ_X . However we will demonstrate that $\rho(\mathbf{r},u)$ has a more complicated structure than $\rho_X^{exact}(\mathbf{r},u)$, thus the approximated exchange hole obtained has a strong oscillatory behavior, making it even more complex than the exact X hole. Subsequently, we apply correlation factors to this X hole model. Despite the success that we achieved recently with our correlation factor models [18–22], applying them to the non-local X-hole models yields a limited accuracy for the prediction of atomization energies, due to the oscillatory behavior of the model X-hole, as we will show below. We believe that we identified one of the problems hampering the development of accurate WDA-type approximations.

In Section 2 we develop an exchange factor such that the resulting X hole satisfies a number of physical conditions, it recovers $\epsilon_X^{exact}(\mathbf{r})$ and it is properly normalized. Building on the notions introduced in Section 2, in Section 3 we provide rigorous lower and upper bounds for $\epsilon_X(\mathbf{r})$ in terms of $\rho(\mathbf{r},u)$. Starting from the approximate non-local X hole, we also model the static correlation in terms of $\rho(\mathbf{r},u)$ (Section 4). These holes are then combined with correlation factors that account for dynamic correlation (Section 5). The various XC functionals are implemented in an electronic structure program (computational details are provided in Section 6) and in Section 7 we assess and analyze the proposed approximations. In Section 8 the conclusions are provided.

2. Non-local approximations of the exchange hole

First, we discuss approximations to the X hole, which we represent through the exchange-factor ansatz. For spin polarized systems we write [19],

$$\rho_X(\mathbf{r},u) = \frac{1 + \zeta(\mathbf{r})}{2} f_X^\sigma(\mathbf{r},u) \rho_\sigma(\mathbf{r},u) + \frac{1 - \zeta(\mathbf{r})}{2} f_X^{\sigma'}(\mathbf{r},u) \rho_{\sigma'}(\mathbf{r},u), \quad (2.1)$$

where $\zeta(\mathbf{r}) = \frac{\rho_\sigma(\mathbf{r}) - \rho_{\sigma'}(\mathbf{r})}{\rho_\sigma(\mathbf{r}) + \rho_{\sigma'}(\mathbf{r})}$ is the spin polarization. The expression that we propose for the exchange factor f_X is original to this work, it is motivated by the strive for simplicity and for ease of the computations necessary to determine the parameters $\alpha, \beta, \gamma, \delta, \lambda$ and η ,

$$f_X^\sigma(\mathbf{r},u) = (\alpha + \beta u^2 + \gamma u^4 + \delta u^6) e^{-\lambda \eta u^2}. \quad (2.2)$$

The parameters are determined at each position in space (e.g., $\alpha = \alpha(\mathbf{r})$), from a number of conditions which we describe now. The known on-top value $\rho(\mathbf{r}, u = 0)$ of the exchange hole requires that

$$f_X^\sigma(\mathbf{r}, u = 0) = -1, \quad (2.3)$$

which yields $\alpha = -1$. β is then chosen to reproduce the exact curvature of the exchange hole $Q_\sigma^{exact}(\mathbf{r})$,

$$\left. \frac{\partial^2 \rho_X^\sigma(\mathbf{r}, u)}{\partial u^2} \right|_{u=0} = -Q_\sigma^{exact}(\mathbf{r}), \quad (2.4)$$

where

$$Q_\sigma^{exact}(\mathbf{r}) = \frac{1}{6} \left(\nabla^2 \rho_\sigma - 2D_\sigma(\mathbf{r}) \right) \quad (2.5)$$

with

$$D_\sigma(\mathbf{r}) = \sum_i |\nabla \phi_{i\sigma}(\mathbf{r})|^2 - \frac{1}{4} \frac{|\nabla \rho_\sigma(\mathbf{r})|^2}{\rho_\sigma(\mathbf{r})}. \quad (2.6)$$

The parameters γ and δ are found by solving the two linear equations,

$$\int_0^\infty 4\pi u^2 \rho_x^\sigma(\mathbf{r}, u) du = -1 \quad (2.7)$$

and

$$\int_0^\infty 2\pi u \rho_X^\sigma(\mathbf{r}, u) du = \epsilon_X^{exact}(\mathbf{r}), \quad (2.8)$$

imposing the normalization of the X-hole and ensuring that the X-hole model reproduces the given exact exchange energy per particle $\epsilon_X^{exact}(\mathbf{r})$. The parameter η is calculated first and sets the range of the X hole. It is determined by equating all the other parameters, except α (which is equal to -1), in $f_X(\mathbf{r}, u)$ to zero and solving,

$$\int_0^\infty 4\pi u^2 \rho_\sigma(\mathbf{r}, u) e^{-\lambda \eta u^2} du = -1. \quad (2.9)$$

For one electron systems, the value of η that satisfies this condition is zero. The same holds for all the other parameters except α which is -1 resulting in $f_X^\sigma(\mathbf{r}, u) = -1$. This ensures that the Coulomb energy is canceled for one-electrons systems so that the model is one-electron SIE free. In some of our models, λ will be used as an empirical global parameter, which will be fitted to minimize the error on atomization energies of a small dataset of 19 molecules, otherwise, it is set to one.

A limitation of Eq. 2.2 is that it does not necessarily satisfy

$$0 \geq \rho_X(\mathbf{r}, u) \geq -\rho(\mathbf{r}, u) \quad (2.10)$$

\Leftrightarrow

$$0 \geq f_X(\mathbf{r}, u) \geq -1 \quad (2.11)$$

for all u . Our approximation to $f_X^\sigma(\mathbf{r}, u)$ can become positive, unlike the exact $f_X^\sigma(\mathbf{r}, u)$, similarly, it can drop below -1 . The ansatz for $f_X^\sigma(\mathbf{r}, u)$ is not flexible enough to satisfy

the conditions Eq. 2.11 in addition to the other constraints. Except for the exponent, the parameters in $f_X^g(\mathbf{r}, u)$ are linear and this greatly facilitates their calculation, in addition to ensuring the existence of solutions. Satisfaction of the constraints Eq. 2.11 requires the introduction of non-linear parameters and then the determination of these parameters becomes numerically difficult and there is no guarantee to obtain a solution or a unique solution. Despite numerous attempts to enforce the constraints Eq. 2.11, we could not find a solution to this problem. Recently, we proposed [29] the use of artificial neural networks to represent X holes and this approach might be suitable to eliminate the problems encountered here with the simple analytic form of $f_X(\mathbf{r}, u)$. The described model for the exchange hole is referred to as f_X^4 and f_X^5 if it uses the empirical parameter λ .

All X hole models presented here reproduce the behavior of the exact X hole under uniform coordinate scaling of the density, where

$$\rho_\Lambda(\mathbf{r}) = \Lambda^3 \rho(\Lambda \mathbf{r}) \quad (2.12)$$

and Λ is a positive scaling factor. For exchange, the exact scaling condition is,

$$\rho_{X\Lambda}(\mathbf{r}, u) = \Lambda^3 \rho_X(\Lambda \mathbf{r}, \Lambda u). \quad (2.13)$$

All our model X holes respect this scaling condition because the conditions that we employ to determine the models have the proper scaling. In more detail, to show that our models respect the exchange scaling condition, we introduce y through $y = k_f u$ with $k_f = (3\pi^2 \rho)^{1/3}$. The X hole is then written as

$$\rho_X(\mathbf{r}, u) = \rho(\mathbf{r}) J_X(\mathbf{r}, y). \quad (2.14)$$

It is easy to verify that $J_X(\mathbf{r}, y)$ is invariant under uniform scaling of the density [30] and the same holds for

$$f_X(\mathbf{r}, y) = \rho(\mathbf{r}) J_X(\mathbf{r}, y) / \rho(\mathbf{r}, y), \quad (2.15)$$

since $\rho(\mathbf{r}, y) / \rho(\mathbf{r})$ is invariant with respect to a uniform scaling. Having established the proper scaling behavior of the exact $f_X(\mathbf{r}, y)$, we note that all the conditions described above to determine the parameters in $f_X(\mathbf{r}, y)$ are invariant with respect to density scaling if formulated in the $y = k_f u$ coordinate, as we showed in our earlier work [31].

The X hole model discussed in this section has a considerable degree of non locality, this is demonstrated in detail in Section 7. Such non-local models require the explicit consideration of static correlation which cancels much of the non-locality of the X hole. Before turning to the inclusion of static and dynamic correlation, we mention a particularly simple exchange hole model that derives from the non-local density.

3. Lower and upper bound to the exchange energy per particle in terms of the non-local electron density

A very simple version of $f_X^\sigma(u)$ is given in terms of the Heavyside step function

$$f_X^\sigma(u) = -\Theta(u_c - u). \quad (3.1)$$

The step function extends up to a certain value of $u = u_c$ where it rises from -1 to 0 . This critical value of u_c is determined by the normalization condition. The corresponding X hole is the negative of $\rho(\mathbf{r}, u)$ or zero. The lower bound for the $\epsilon_X(\mathbf{r})$ is realized by the maximally localized $f_X^\sigma(u)$ of Eq. 3.1 which yields

$$\epsilon_{X,\sigma}^{\text{LB}} = -2\pi \int_0^{u_c} du u \rho_\sigma(r, u). \quad (3.2)$$

While the strong correlation limit of density functional theory [32] provides a lower bound to the XC energy, Eq. 3.2 gives such a bound for $\epsilon_X(\mathbf{r})$,

$$\epsilon_{X,\sigma}(\mathbf{r}) \geq \epsilon_{X,\sigma}^{\text{LB}}(\mathbf{r}), \quad (3.3)$$

where the inequality turns into an equality for one-electron systems.

To prove the inequality 3.2, we minimize the $\epsilon_X(\mathbf{r})$ with respect to variations of $f_X^\sigma(u)$, subject to the constraints of X-hole normalization and

$$0 \geq f_X^\sigma(u) \geq -1. \quad (3.4)$$

The inequalities 3.4 can be satisfied by introducing the function $g_X^\sigma(u)$ such that

$$f_X^\sigma(u) = -\sin^2[g_X^\sigma(u)]. \quad (3.5)$$

We then use the Lagrangian multiplier method to minimize $\epsilon_{X,\sigma}(\mathbf{r})$ with respect to $g_X^\sigma(u)$, including the constraint of normalization, which yields the Lagrangian

$$L[g_X^\sigma] = 2\pi \int du u \sin^2[g_X^\sigma(u)] \rho_\sigma(\mathbf{r}, u) + \lambda \left[4\pi \int du u^2 \sin^2[g_X^\sigma(u)] \rho_\sigma(\mathbf{r}, u) - 1 \right]. \quad (3.6)$$

The stationary points (with respect to $g_X^\sigma(u)$ and λ) of this functional satisfy

$$2 \pi u \rho_\sigma(\mathbf{r}, u) \sin[g_X^\sigma(u)] \cos[g_X^\sigma(u)] + \lambda 4\pi u^2 \rho_\sigma(\mathbf{r}, u) \sin[g_X^\sigma(u)] \cos[g_X^\sigma(u)] = 0 \quad (3.7)$$

\Leftrightarrow

$$(1 + 2\lambda u) \sin[g_X^\sigma(u)] \cos[g_X^\sigma(u)] = 0 \quad (3.8)$$

and

$$4\pi \int du u^2 \sin^2[g_X^\sigma(u)] \rho_\sigma(\mathbf{r}, u) - 1 = 0. \quad (3.9)$$

The stationary solutions of Eq. 3.8 are provided by $g_X^\sigma(u) = \frac{n\pi}{2}$, n being an integer, or equivalently $f_X^\sigma(u) = 0, -1$. To ensure normalization of the X hole, both solutions need to be combined. According to the theory of Lagrangian multipliers, variation of λ permits the

adjustment of the normalization integral to recover the correct value of -1 . A certain value of λ corresponds to a critical value of u referred to as u_c which ensures that Eq. 3.8 vanishes

$$\lambda = -\frac{1}{2u_c}. \quad (3.10)$$

We identify this value of u_c as the critical radius where the solution $f_X^\sigma(u) = -\sin^2[g_X^\sigma(u)] = -1$ switches over to $f_X^\sigma(u) = -\sin^2[g_X^\sigma(u)] = 0$, so at this particular point where $-\sin^2[g_X^\sigma(u)]$ jumps from -1 to 0 , the factor $1 + 2\lambda u_c$ ensures that the derivative Eq. 3.8 vanishes. The employed Lagrangian multiplier method identifies the extrema of Eq. 3.6. The only other extremum (maximum) of $L[g_X^\sigma]$ is provided by $f_X^\sigma(u) = -\Theta(u - u_c)$, where initially the X hole is zero and at $u = u_c$ is switches to the negative of $\rho_\sigma(\mathbf{r}, u)$. This solution provides a rigorous upper bound to $\epsilon_{X,\sigma}(\mathbf{r})$. Since the ratio of the integrands of $\epsilon_{X,\sigma}$ and the normalization integral, i.e., $\frac{2\pi u f_X^\sigma(u) \rho_\sigma(\mathbf{r}, u)}{4\pi u^2 f_X^\sigma(u) \rho_\sigma(\mathbf{r}, u)}$ is strictly decreasing, the solution Eq. 3.1 is only compatible with the minimization of the expression Eq. 3.6, it is advantageous to have the deepest possible hole close to the reference point \mathbf{r} rather than moving the hole away from \mathbf{r} . For a one-electron system, the described upper and lower bound to $\epsilon_{X,\sigma}(\mathbf{r})$ coincide.

4. Modeling static correlation within the non-local approach

The correlation energy is often divided into dynamic and static contributions which require different approximation strategies. Dynamic correlation is a short-range contribution to the XC hole, while static correlation is a long range feature that is difficult to approximate within DFT. Here we consider two approaches to model static correlation within the non-local ansatz, to the best of our knowledge, there are no previous attempts to model static correlation starting from $\rho(\mathbf{r}, u)$. The first model is inspired by the work of Becke [33], and we refer to this fully non-local approximation as opposite-spin static (OS) correlation energy functional. For the second, we employ a multi-configurational (MC) reference state, following our previous work on the correlation factor model [22].

In the first method, we consider static correlation through exclusion of the opposite-spin electron from the vicinity of the reference electron [33],

$$\rho_{XS,\sigma}^{OS}(\mathbf{r}, u) = \rho_{X,\sigma}(\mathbf{r}, u) + a_{\sigma'} \rho_{X,\sigma'}(\mathbf{r}, u). \quad (4.1)$$

The resulting $\rho_{XS,\sigma}^{OS}(\mathbf{r}, u)$ are then combined using the conventional formula for the spin average X hole to yield the exchange plus static correlation (XS) hole,

$$\rho_{XS}^{OS}(\mathbf{r}, u) = \left(\frac{(1 + \zeta)}{2} \rho_{XS,\sigma}^{OS}(\mathbf{r}, u) + \frac{(1 - \zeta)}{2} \rho_{XS,\sigma'}^{OS}(\mathbf{r}, u) \right) e^{-\kappa u^2} \quad (4.2)$$

where κ is introduced to recover the normalization, $4\pi \int_0^\infty du u^2 \rho_{\text{XS}}^{\text{OS}}(\mathbf{r}, u) = -1$, which is lost for $\rho_{\text{XS},\sigma}^{\text{OS}}(\mathbf{r}, u)$ because of the addition of an opposite spin component in Eq. 4.1. The effect of adding an opposite-spin component X hole in Eq. 4.1 is to increase the depth of the hole at short u values; consequently, the range of the resulting spin-averaged hole has to be shortened through $e^{-\kappa u^2}$ to ensure normalization.

To determine the parameters $a_\sigma, a_{\sigma'}$, we consider the normalization (N_σ^{BR}) of the reverse Becke-Roussel (BR) exchange hole construction, reproducing the exact exchange energy per particle [33, 34]. The conventional BR hole yields an exchange energy per particle $\epsilon_{\text{X},\sigma}^{\text{BR}}(\mathbf{r})$ for a certain value of ρ and the curvature Q as input while respecting the normalization. However, the reverse BR hole returns a normalization when using the exact $\epsilon_{\text{X},\sigma}(\mathbf{r})$ as the third input. The value of the normalization represents the fraction of the hole localized on an atom. If the exact exchange hole is localized on an atomic scale, the effective normalization is close to -1 . This will lead to a value of $a_{\sigma'} = 0$ where the static correlation is negligible. However, if the exact exchange hole is quite delocalized, the BR hole normalization will be above -1 , indicating the presence of static correlation. In this case, we need a positive, nonzero $a_{\sigma'}$. Following Becke [33] we determine the parameters a_σ and $a_{\sigma'}$ from N_σ^{BR} through,

$$a_\sigma = a_{\sigma'} = a \quad (4.3)$$

$$a = \min \left((1 - N_\sigma^{\text{BR}})/N_{\sigma'}^{\text{BR}}, (1 - N_{\sigma'}^{\text{BR}})/N_\sigma^{\text{BR}}, 1 \right), \quad (4.4)$$

where N_σ^{BR} represents the number of electrons (normalization) of the reverse BR hole procedure for a spin channel.

In the second method to account for static correlation, a MC reference state is used to determine an exchange plus static correlation (XS) hole. The XC hole at full coupling strength of a given wave function is defined in terms of the spin-averaged pair density ($P^{\text{MC}}(\mathbf{r}, u)$) by,

$$\rho_{\text{XC}}^{\text{MC}}(\mathbf{r}, u) = \frac{P^{\text{MC}}(\mathbf{r}, u)}{\rho(\mathbf{r})} - \rho(\mathbf{r}, u). \quad (4.5)$$

Instead of using this hole directly, we replace it by an exchange-type model hole which reproduces the same XS energy as $\rho_{\text{XC}}^{\text{MC}}(\mathbf{r}, u)$,

$$\rho_{\text{XS}}^{\text{MC}}(\mathbf{r}, u) = \frac{1 + \zeta(\mathbf{r})}{2} \rho_{\text{XS},\sigma}^{\text{MC}}(\mathbf{r}, u) + \frac{1 - \zeta(\mathbf{r})}{2} \rho_{\text{XS},\sigma'}^{\text{MC}}(\mathbf{r}, u). \quad (4.6)$$

As in our previous CF models, we impose the condition that the on-top value ($u = 0$) of the model hole reproduces the one of the MC XS hole $\rho_{\text{XS}}^{\text{MC}}(\mathbf{r}, 0)$,

$$\rho_{\text{XC}}^{\text{MC}}(\mathbf{r}, 0) = \rho_{\text{XS}}^{\text{MC}}(\mathbf{r}, 0) \quad (4.7)$$

$$\begin{aligned} &= \frac{1 + \zeta(\mathbf{r})}{2} \rho_{\text{XS},\sigma}^{\text{MC}}(\mathbf{r}, 0) + \frac{1 - \zeta(\mathbf{r})}{2} \rho_{\text{XS},\sigma'}^{\text{MC}}(\mathbf{r}, 0) \\ &= \frac{1 + \zeta^2(\mathbf{r})}{2} \rho(\mathbf{r}). \end{aligned} \quad (4.8)$$

To ensure that the latter equation can be satisfied, the effective spin polarization ($\zeta^*(\mathbf{r})$) is introduced which is determined such that Eq. 4.8 is satisfied [22, 35–37],

$$\zeta^*(\mathbf{r}) = \sqrt{\frac{2\rho_{\text{XS}}^{\text{MC}}(\mathbf{r}, u=0)}{\rho(\mathbf{r})} - 1} \quad (4.9)$$

We then define $\rho_{\text{XS},\sigma(\sigma')}^{\text{MC}}(\mathbf{r}, u)$ as

$$\rho_{\text{XS},\sigma(\sigma')}^{\text{MC}}(\mathbf{r}, u) = \frac{1 + (-)\zeta^*(\mathbf{r})}{2} \rho_{\text{X}}^{\text{NL}}(\mathbf{r}, u). \quad (4.10)$$

The $\rho_{\text{X}}^{\text{NL}}$ hole is the one of Section 2, however we replace the curvature $Q_{\sigma}^{\text{exact}}$ of Eq. 2.4 and the energy density $\epsilon_{\text{X}}^{\text{exact}}$ of Eq. 2.8 by their MCSCF (Multi-configurational Self-Consistent-Field) values, see the Appendix for more details on the calculations.. The model using this procedure will be named $f_{\text{XS}}^{\text{MC}}$, while those using the OS approach with the X hole of Section 2 $f_{\text{XS}}^{4\text{OS}}$ or $f_{\text{XS}}^{5\text{OS}}$ if they use the empirical parameter.

5. Modeling dynamical correlation

The correlation factor ($f_{\text{C}}(\mathbf{r}, u)$) ansatz [18–22, 38] which we developed recently is now employed to account for dynamic correlation effects,

$$\rho_{\text{XC}}(\mathbf{r}, u) = f_{\text{C}}(\mathbf{r}, u) \rho_{\text{XS}}(\mathbf{r}, u). \quad (5.1)$$

There are two versions of $f_{\text{C}}(\mathbf{r}, u)$ that we consider: a simple three parameter model f_{C}^3 and the CFX correlation factor [18–22] published previously. Both of these factors are now briefly described.

5.1. The three parameter $f_{\text{C}}^3(\mathbf{r}, u)$ correlation factor

In previous work, we developed a rather sophisticated correlation factor that we describe below. However, first we present what is maybe the simplest version of the correlation factor which depends on three parameters,

$$f_{\text{C}}^3(\mathbf{r}, u) = (A + Cu^2)e^{-Bu}. \quad (5.2)$$

The \mathbf{r} dependence of the parameters is omitted for simplicity of the notation. In this version, A is adjusted to reproduce the Local Spin Density approximation (LSD) on-top value of the XC hole [19, 30],

$$\rho_{\text{XC}}(\mathbf{r}, u=0) = \rho_{\text{XC}}^{\text{LSD}}(\mathbf{r}, u=0) \quad (5.3)$$

B is determined through the cusp condition [19, 30],

$$\left. \frac{\partial \rho_{\text{XC}}(\mathbf{r}, u)}{\partial u} \right|_{u=0} = c^{\text{LSD}}(\mathbf{r}). \quad (5.4)$$

The details for the on-top and cusp conditions can be found in references[19, 30]. Finally, C enforces the normalization condition of the XC hole and this leads to,

$$C = -\frac{1/4\pi + A \int_0^\infty u^2 \rho_X(\mathbf{r}, u) e^{-Bu} du}{\int_0^\infty u^4 \rho_X(\mathbf{r}, u) e^{-Bu} du}. \quad (5.5)$$

5.2. Reviewing the CFX correlation factor

We also adapt our recent version of the CF model to account for dynamic correlation within the non-local model. The CF of reference [20] is

$$f_C^{\text{CFX}}(\mathbf{r}, u) = (A + Bu + Cu^2 + (1 - \Sigma)Du^4)e^{-(1-\Sigma)\mu u^2}. \quad (5.6)$$

The six parameters of $f_C^{\text{CFX}}(\mathbf{r}, u)$ are again functions of \mathbf{r} and they are determined through a number of conditions imposed on the XC-hole $\rho_{\text{XC}}(\mathbf{r}, u)$. The parameters A and B ensure that: (i) the depth of the on-top hole ($\rho_{\text{XC}}(\mathbf{r}, u = 0)$) is determined through the local approximation [18, 19] and (ii) $\rho_{\text{XC}}(\mathbf{r}, u)$ satisfies the Coulomb cusp condition [30] analogous to the f_C^3 model. In this model, B is a linear parameter instead of being in the exponential. To determine the parameters (μ, D) , we require that the CF model for $\rho_{\text{XC}}(\mathbf{r}, u)$ reproduces the LSD and PBE XC energy per particle, given that the underlying X-hole reproduces the LSD and PBE X energy per particle,

$$\int d^3u \frac{\rho_{\text{XC}}(\mathbf{r}, u)}{2u} = \epsilon_{\text{XC}}^y(\mathbf{r}), \quad (5.7)$$

where y stands for LSD or PBE (for further details, we refer to our previous work on the CF model [18–20, 20, 22]). Since neither LSD nor PBE are one-electron, self-correlation free, this error is built into the coefficients μ and D and to correct the self-correlation error, μ and D are scaled by $(1 - \Sigma)$, a term which vanishes in one-electron regions and ensures that the CF reduces to one there. The coefficient Σ that we employ is the one that we proposed previously in Ref. [20] where we referred to the corresponding functional as CFX Σ_2 . The parameters are obtained using the local BR X-hole, since no solutions were found by root finding algorithm for the parameter μ . This model has a flexible parameter a which can be fitted to data, its value will be given in the table caption (Tables II-4 to II-7). Finally, the last remaining parameter C is determined through the normalization condition imposed on the resulting XC-hole which we refer to by $\rho_{\text{XC}}^{\text{CFXS}}(\mathbf{r}, u) = f_C(\mathbf{r}, u)\rho_{\text{XS}}(\mathbf{r}, u)$,

$$\int d^3u \rho_{\text{XC}}^{\text{CFXS}}(\mathbf{r}, u) = -1. \quad (5.8)$$

Unlike our earlier CFX model [18–22], the intermediate XS hole is normalized and this ensures that the resulting XC model is free of the SIE. The model using this correlation factor will be referred as f_C^{CFX} .

5.3. Avoiding double-counting of correlation effects

To ensure that the CF for dynamic correlation (both f_C^{CFX} and f_C^3) is compatible with the XS hole and double-counting of correlation effects is minimized, we employ an effective $\zeta^*(\mathbf{r})$ in the evaluation of the dynamic correlation factors. As reviewed above, a key ingredient of the CF is the on-top value of the XC hole, the amount of correlation energy introduced through the CF strongly depends on it. The on-top value, in turn, is dependent on $\zeta(\mathbf{r})$ and to match the CF to the XS hole, the same spin-polarization should be used for both functions. Employing $\zeta^*(\mathbf{r})$ of equation 4.9 (for the OS approach, we replace $\rho_{\text{XS}}^{\text{MC}}$ by $\rho_{\text{XS}}^{\text{OS}}$) in the CF ensures that the on-top value of the XC-hole is that of a physical system and respects the condition $-\rho(\mathbf{r}) \leq \rho_{\text{XC}}(\mathbf{r},0)$.

6. Computational details

The models described above are implemented within the PySCF program [39, 40]. Using the various ingredients provided by PySCF (electronic density, density gradient, etc.), a Python and C++ code has been developed for our models. The integration over the inter-electronic distance u is realized with a uniform grid of 10 000 points from 0 to 30 Bohr. The molecular structures are optimized at the PBE/6-311+g(2d,p) level. The computations are performed non-self-consistently using self-consistent PBE/6-311+g(2d,p) or MCSCF/6-311+g(2d,p) orbitals depending on whether the model uses exact exchange or MCSCF energy densities. For the MCSCF calculation, the complete active spaces consists of the valence orbitals and electrons. The details about the calculation of $\rho(\mathbf{r},u)$ for gaussian basis sets can be found in references [16, 27], while for the exact exchange hole it can be found in [29]. In the Appendix, we explain how the curvature and the on-top value of the MCSCF XC hole are calculated.

For the calculation of the reverse BR exchange hole normalization [33], the lower limit was set to -1. This is necessary because if the normalization drops below -1, then ζ^* would be imaginary.

7. Assessing the non-local functionals

First, we review the X and XC functionals that emerge as combination of the various exchange and correlation factors introduced in the sections Secs. 2, 4, and 5. For the X hole, we introduced the four parameter model f_X^4 and in addition, we modified this model with one additional empirical parameter (λ) to yield f_X^5 . Both factors are described in Sections 2. The λ empirical parameter is obtained by minimizing the mean squared error of the atomization energy dataset of 19 molecules. For the static correlation, we designed two models, the OS approach without ($f_{\text{XS}}^{4\text{OS}}$) and with additional empirical parameter ($f_{\text{XS}}^{5\text{OS}}$) as

Tab. II-1. Exchange and exchange-correlation factors developed in this work. f_X^4 and f_X^5 represent the basic exchange factors, where f_X^5 contains one additional empirical parameter (λ) compared to f_X^4 . Both of these models are described in Section 2. To account for static correlation (XS denotes exchange plus static correlation), the opposite-spin static correlation (OS) and Multi-Configurational reference (MC) approaches are described in Section 4. Dynamic correlation is accounted for by two correlation factors (f_C^3 and f_C^{CFX}) which are detailed in Section 5.1 and 5.2, respectively. The various exchange and correlation factors are then combined as indicated in the table and applied to $\rho(\mathbf{r},u)$ to obtain XC functionals.

Correlation	Factors reproducing exact exchange	Factors drawing on MCSCF
Static	$f_{\text{XS}}^{4\text{OS}}, f_{\text{XS}}^{5\text{OS}}$	$f_{\text{XS}}^{4\text{MC}}$
Dynamic	$f_C^{\text{CFX}} f_X^4, f_C^3 f_X^4$	
Static+Dynamic	$f_C^{\text{CFX}} f_{\text{XS}}^{5\text{OS}}, f_C^{\text{CFX}} f_{\text{XS}}^{4\text{OS}}, f_C^3 f_{\text{XS}}^{5\text{OS}}, f_C^3 f_{\text{XS}}^{4\text{OS}}$	$f_C^{\text{CFX}} f_{\text{XS}}^{4\text{MC}}, f_C^3 f_{\text{XS}}^{4\text{MC}}$

well as the MCSCF based factor ($f_{\text{XS}}^{\text{MC}}$), both described in Section 4. Furthermore, we employ two correlation factors, the previously developed f_C^{CFX} which is reviewed in Section 5.2 and f_C^3 constructed in Section 5.1. Combining these approximations for exchange, static, and dynamic correlation yields various functionals that we assess (see table II-1 for a summary): $f_C^{\text{CFX}} f_{\text{XS}}^{5\text{OS}}, f_C^{\text{CFX}} f_{\text{XS}}^{4\text{OS}}, f_C^3 f_{\text{XS}}^{5\text{OS}}, f_C^3 f_{\text{XS}}^{4\text{OS}}, f_C^{\text{CFX}} f_X^4, f_C^3 f_X^4, f_C^{\text{CFX}} f_{\text{XS}}^{4\text{MC}}, f_C^3 f_{\text{XS}}^{4\text{MC}}, \text{CFX}, \text{CFXS}$ and PBE. The latter three functionals have been added as references for comparison purposes. The first functional CFX is the correlation factor model of Ref. [21] that builds on a single reference determinant, CFXS [22] exploits multi-configurational reference states and PBE is the Perdew-Burke-Ernzerhof functional [41].

As explained in the introduction, one motivation for the present work is to examine the non-local approach for the construction of X-hole models. In particular, we are interested in using these models in subsequent attempts to add correlation. One element which distinguishes our approach from the previous works is that the reproduction of the exact exchange energy per particle is imposed as a condition on the X-hole model. This is in line with the other recent attempts [33, 42–44] to build XC functionals starting from exact exchange.

The properties that we focus on to assess the various models are the total energies of atoms and the atomization energies of a set of molecules. The total energies are useful to ensure that the functional errors don't scale dis-proportionally with atom size. The second property, the atomization energy, is a crucial property for chemistry and it represents a severe test for approximate functionals, they have to provide a balanced account of the static and dynamic correlation contributions.

7.1. Accounting for exchange and static correlation

The results for the atomic and atomization energies for the static correlation models $f_{\text{XS}}^{5\text{OS}}$ and $f_{\text{XS}}^{4\text{OS}}$, described in Section 4, are shown in Tables II-2 and II-3 respectively. As

Tab. II-2. Total energies for atoms in Hartree units, MAE abbreviates Mean Absolute Error. The approximations considered are characterized by their factors f as described in Table II-1. For $f_{\text{XS}}^{5\text{OS}}$, $\lambda = 0.82$.

Atom	$f_{\text{XS}}^{5\text{OS}}$	$f_{\text{XS}}^{4\text{OS}}$	E_{X}	MCSCF	exp. [45–47]
H	-0.499	-0.499	-0.499	-0.500	-0.500
He	-2.873	-2.873	-2.859	-2.860	-2.904
Li	-7.446	-7.445	-7.431	-7.432	-7.478
Be	-14.592	-14.589	-14.570	-14.572	-14.667
B	-24.553	-24.549	-24.528	-24.562	-24.654
C	-37.714	-37.710	-37.686	-37.704	-37.845
N	-54.427	-54.422	-54.394	-54.395	-54.589
O	-74.859	-74.852	-74.802	-74.802	-75.067
F	-99.482	-99.472	-99.394	-99.396	-99.734
Ne	-128.649	-128.636	-128.519	-128.527	-128.938
Na	-161.966	-161.953	-161.839	-161.846	-162.255
Mg	-199.731	-199.717	-199.600	-199.607	-200.053
Al	-241.997	-241.983	-241.865	-241.894	-242.346
Si	-288.974	-288.961	-288.843	-288.861	-289.359
P	-340.833	-340.819	-340.702	-340.709	-341.259
S	-397.629	-397.615	-397.491	-397.496	-398.110
Cl	-459.611	-459.597	-459.467	-459.470	-460.148
Ar	-526.952	-526.937	-526.800	-526.807	-527.540
MAE	0.259	0.268	0.342	0.334	-

expected, since the X-holes for atoms are already localized, MCSCF does not add significant correlation energies compared to exact exchange (E_{X}) for atoms. Note that E_{X} is calculated from the PBE self-consistent orbitals. $f_{\text{XS}}^{4\text{OS}}$ model adds too much static correlation compared to MCSCF, we note that by construction, it is exact for the hydrogen atom. The empirical parameter λ in $f_{\text{XS}}^{5\text{OS}}$ increases the absolute value of the correlation energy slightly. With a λ value below 1, the XS-hole is less localized, thus we would expect less correlation. However, as we will show by analyzing the underlying hole models, the behavior of f_x^4 is difficult to interpret and predict due to the oscillating behavior of $\rho(\mathbf{r}, u)$. For the atomization energies of molecules with a single bonds, the MAE of $f_{\text{XS}}^{4\text{OS}}$ (35 kcal/mol) is relatively close to MCSCF(29 kcal/mol). Both underestimate the atomization energies since dynamic correlation is missing, while E_{X} without any correlation predicts even lower atomization energies. For molecules with multiple bonds, the MAE for MCSCF (29kcal/mol) is almost the same as for those with a single bonds, while it’s much worse for E_{X} . Thus, MCSCF adds significantly more static correlation for multiple bonds, as expected. $f_{\text{XS}}^{5\text{OS}}$ and $f_{\text{XS}}^{4\text{OS}}$ do not add enough static correlation compared to MCSCF, but significantly improve upon E_{X} . As Becke observed [33], the O_2 molecules which has a triplet state is poorly represented by the OS model which does not include static correlation for parallel spin. However, we also obtain

Tab. II-3. Atomization energies in kcal/mol for a set of single and multi-bonded molecules. MAE and ME are the mean absolute error and mean error, respectively. For $f_{\text{XS}}^{5\text{OS}}$: $\lambda = 0.82$.

Molecule	$f_{\text{XS}}^{5\text{OS}}$	$f_{\text{XS}}^{4\text{OS}}$	E_{X}	MCSCF	exp. [48–50]
H ₂	86	86	83	95	109
LiH	36	35	33	44	58
CH ₄	349	346	325	370	419
NH ₃	234	230	197	248	297
H ₂ O	188	185	152	192	232
HF	119	116	95	114	141
Li ₂	6	3	3	10	24
LiF	115	111	88	94	139
Be ₂	-3	-6	-12	-8	3
MAE	32.4	35.0	50.7	29.3	0
ME	-32.4	-35.0	-50.7	-29.3	0
C ₂ H ₂	344	336	287	357	405
C ₂ H ₄	474	467	421	499	563
HCN	261	253	191	285	312
CO	230	223	167	251	259
N ₂	185	176	107	211	229
NO	102	95	42	128	153
O ₂	57	52	21	90	121
F ₂	12	5	-45	16	39
P ₂	59	55	28	95	117
Cl ₂	30	27	14	38	58
MAE	50.1	56.9	102.3	28.6	0
ME	-50.1	-56.9	-102.3	-28.6	0
Tot. MAE	41.7	46.5	77.9	28.9	0
Tot. ME	-41.7	-46.5	-77.9	-28.9	0

poor results for other molecule such as P₂. Further investigation is required to confirm if it’s also related to static correlation of parallel spin in multiple bonds. The addition of the empirical λ parameter improves the results slightly but does not change the overall trends.

For the N₂ molecules, the model holes are investigated to gain a better understanding of our results. In Figure 2.1 we compare our models with the exact X-hole, the hybrid BR X hole of CFX [19], the X holes without static correlation (f_x^4) and $\rho(\mathbf{r}, u)$.

For $z = 1$ (z is the position in Bohr along the molecular axis), which is between the two N atoms, f_{X}^4 does oscillate, but it’s a good approximation to the exact X hole. The CFX hybrid X hole is slightly more long ranged, but still close to the exact X hole. For the core region of an atom ($z = 0$), at short u values, all the approximations are fairly good. However, for f_{X}^4 , the negativity constraint is not satisfied for all values of u . Where our non-local models fail and the oscillation gets significant is for the core-valence region ($z = \pm 0.6$). The X hole of the CFX model is quite close to the exact one. Similarly, other researchers also found that

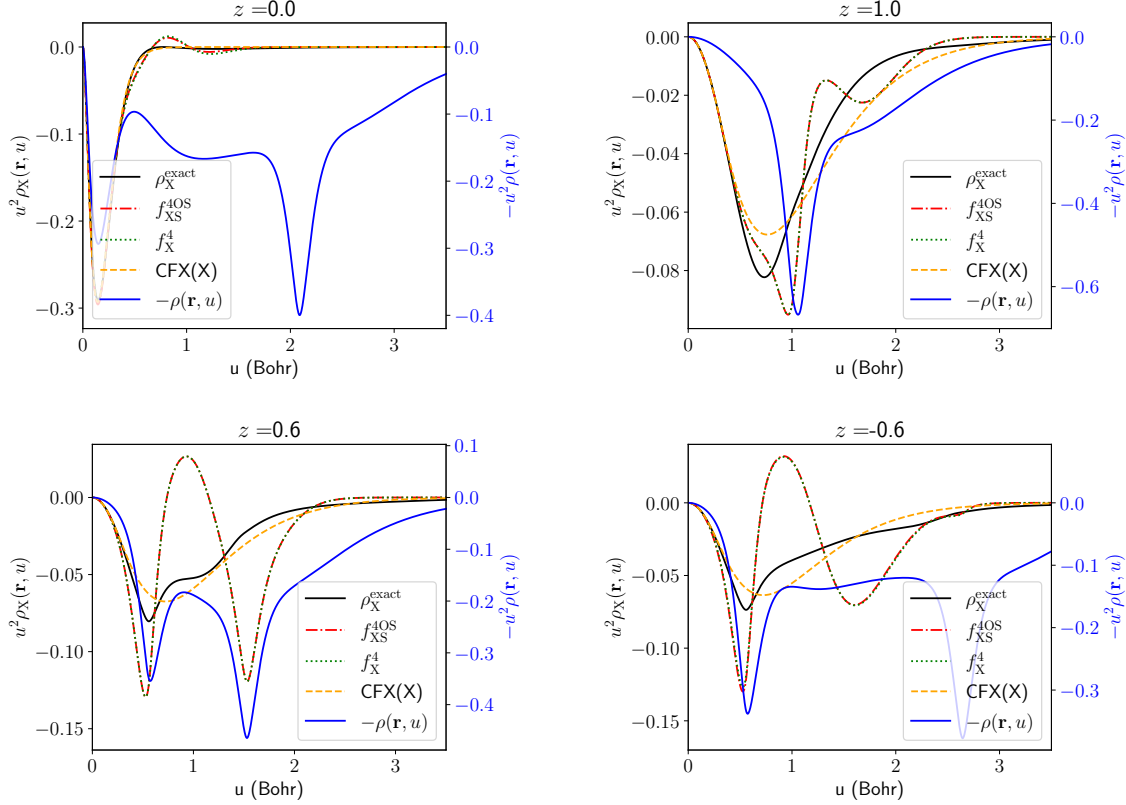


Fig. 2.1. Comparison of exchange and exchange-plus-static-correlation holes for the N_2 molecule. One N atom is at the position (0.00,0.00,0.00) and the other at (0.00,0.00,2.08) in Bohr. z is the position along the molecular axis. CFX(X) indicates that the X hole of the CFX method is represented, the other approximations are identified with their factor f as described in Table II-1.

the BR hole is an excellent approximation to the exact X hole [51]. The f_X^4 exchange factor does not suppress the oscillation of $\rho(\mathbf{r}, u)$ enough, the BR model which does not consider the non-local density does not need to suppress the oscillations.

Figure 2.2 displays the exact exchange factor for the N_2 molecule ($f_X^{exact}(\mathbf{r}, u) = \rho_X^{exact}(\mathbf{r}, u)/\rho(\mathbf{r}, u)$) and compares it to the $f_X^4(\mathbf{r}, u)$ model. While our models are good at short u values, they deteriorate for larger u values, particularly for reference positions \mathbf{r} in the core-valence region. As discussed in Section 2, in relation to the inequality 2.11, $f_X^4(\mathbf{r}, u)$ can become positive, and it does in small regions close to the core. However, in the cases shown in the Figure 2.2, $f_X^4(\mathbf{r}, u) \geq -1$ is respected even though it is not explicitly imposed.

7.2. Dynamic correlation with correlation factors

To add dynamic correlation, we consider the f_C^3 and f_C^{CFX} correlation factors described in Section 5.1 and 5.2, respectively. For atomic energies, the results are shown in Table II-4. Most models have a MAE below 0.1 Hartree, except $f_C^3 f_X^4$. This model does not include

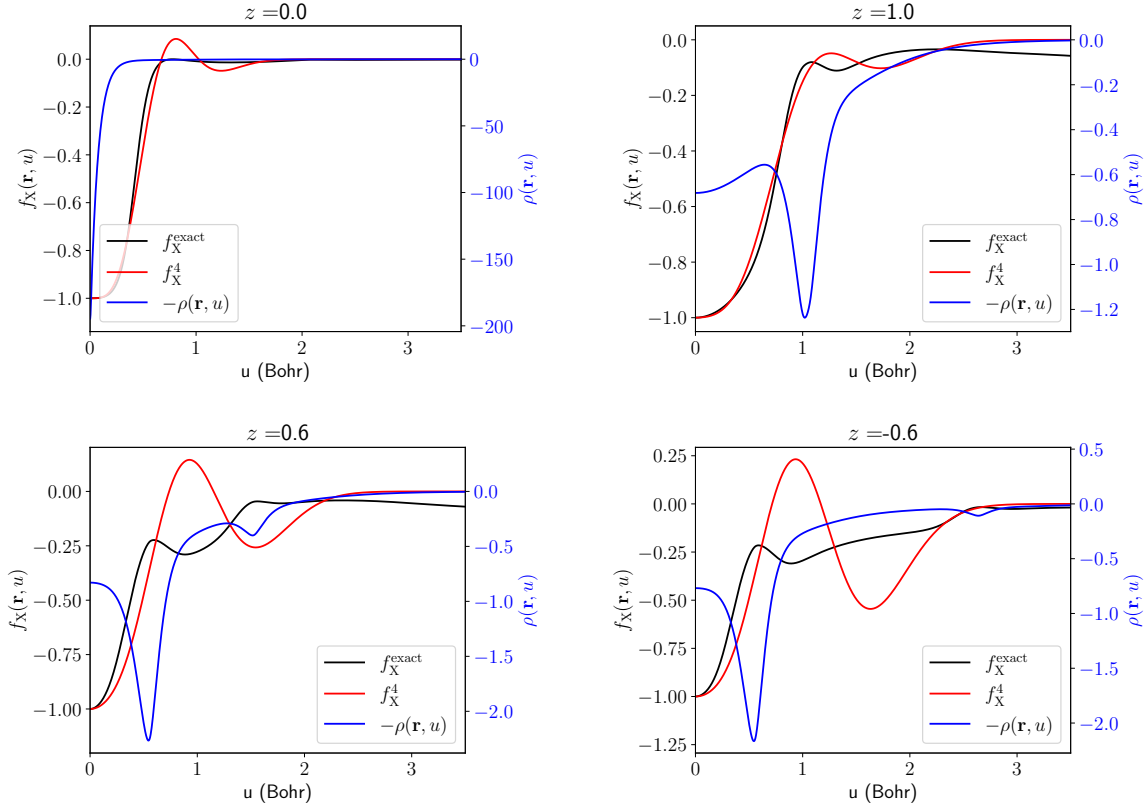


Fig. 2.2. Comparison of exchange factors for the N_2 molecule. One N atom is at the position $(0.00, 0.00, 0.00)$ and the other at $(0.00, 0.00, 2.08)$ in Bohr. z is the position along the molecular axis.

static correlation. While static correlation should be negligible for atoms, we see in Table II-2 that the OS method adds some for our model, which explains the improvement for $f_C^3 f_{XS}^{4OS}$. The MAE is particularly small for $f_C^{CFX} f_{XS}^{4OS}$, but in general most model are satisfactory for atoms when compared to CFX or PBE.

The atomization energies are shown in Table II-5. For both the f_C^{CFX} and f_C^3 correlation factors, we observe that the addition of static correlation through the OS approach can significantly improves the results as revealed by comparing $f_C^{CFX} f_X^4$ and $f_C^3 f_X^4$ vs $f_C^{CFX} f_{XS}^{4OS}$ and $f_C^3 f_{XS}^{4OS}$. Since for f_C^3 the atomization energies are systematically underestimated, the MAE drops upon addition of static correlation, unlike for f_C^{CFX} , where it increases. We show below that for N_2 , f_C^{CFX} amplifies the spurious oscillation of the XS holes. The parameters for the f_C^{CFX} correlation factors are developed for a local hole (the BR X-hole) and are not adapted for a non-local hole. f_C^3 only applies a fairly local modification, thus has a smaller propensity for spurious oscillations. Furthermore, as mentioned in the previous section, we observe worse results for molecules with multiple bonds and O_2 . To the best of our knowledge,

Tab. II-4. Total energies for atoms in Hartree and the Mean Absolute Error (MAE) for different methods. For $f_C^{\text{CFX}} f_{\text{XS}}^{\text{OS}}$, $a = 0.13$ and $\lambda = 1.30$. For $f_C^3 f_{\text{XS}}^{\text{OS}}$, $\lambda = 0.82$

Atom	$f_C^{\text{CFX}} f_{\text{XS}}^{5\text{OS}}$	$f_C^{\text{CFX}} f_{\text{XS}}^{4\text{OS}}$	$f_C^3 f_{\text{XS}}^{5\text{OS}}$	$f_C^3 f_{\text{XS}}^{4\text{OS}}$	$f_C^{\text{CFX}} f_X^4$	$f_C^3 f_X^4$	CFX	PBE	exp.
H	-0.499	-0.499	-0.499	-0.499	-0.499	-0.499	-0.500	-0.500	-0.500
He	-2.913	-2.911	-2.915	-2.915	-2.902	-2.903	-2.911	-2.891	-2.904
Li	-7.483	-7.478	-7.489	-7.489	-7.467	-7.475	-7.485	-7.461	-7.478
Be	-14.648	-14.646	-14.675	-14.671	-14.631	-14.654	-14.661	-14.628	-14.667
B	-24.622	-24.626	-24.650	-24.646	-24.610	-24.625	-24.638	-24.609	-24.654
C	-37.802	-37.815	-37.818	-37.815	-37.798	-37.791	-37.821	-37.795	-37.845
N	-54.534	-54.558	-54.533	-54.531	-54.537	-54.503	-54.559	-54.530	-54.589
O	-75.017	-75.065	-75.005	-75.000	-75.027	-74.950	-75.033	-75.005	-75.067
F	-99.682	-99.752	-99.663	-99.653	-99.688	-99.576	-99.688	-99.661	-99.734
Ne	-128.883	-128.973	-128.857	-128.843	-128.875	-128.730	-128.876	-128.847	-128.938
Na	-162.213	-162.310	-162.177	-162.163	-162.213	-162.052	-162.181	-162.158	-162.255
Mg	-199.994	-200.102	-199.963	-199.944	-200.002	-199.830	-199.967	-199.945	-200.053
Al	-242.277	-242.391	-242.239	-242.220	-242.290	-242.106	-242.234	-242.228	-242.346
Si	-289.270	-289.391	-289.226	-289.208	-289.289	-289.093	-289.216	-289.226	-289.359
P	-341.145	-341.272	-341.092	-341.075	-341.170	-340.960	-341.081	-341.106	-341.259
S	-397.975	-398.111	-397.926	-397.905	-398.003	-397.785	-397.901	-397.940	-398.110
Cl	-459.990	-460.136	-459.942	-459.919	-460.023	-459.794	-459.903	-459.960	-460.148
Ar	-527.362	-527.518	-527.313	-527.288	-527.398	-527.157	-527.258	-527.332	-527.540
MAE	0.065	0.022	0.085	0.095	0.057	0.165	0.087	0.090	-

while the results obtained with $f_C^3 f_{\text{XS}}^{5\text{OS}}$ and $f_C^3 f_{\text{XS}}^{4\text{OS}}$ are not yet as good as those obtained with functionals such as CFX [19], they are not surpassed by other non-local functionals.

Finally, we compare XC-holes of the various models to the XC-hole obtained with the CFX approach. Since the CFX yields the most accurate results for atomization energies, we use it as a reference point and consider it to be the most accurate approximation that we have presently at our disposition. In the core region, the various models yield similar XC-holes. Notably, all the XC holes based on the non-local density show a more oscillatory behavior compared to the CFX model. This tendency is even stronger in the valence region at $z = 1.0$ where the oscillations already observed in the X-hole carry over to the XC-hole models. At $z = \pm 0.6$ the oscillations are extremely large, even though the model holes seem to oscillate around the CFX hole. The diminished performance of the non-local models compared to CFX can be explained based on the analysis presented. The non-local holes tend to have a highly oscillatory behavior and any correlation factor model that is applied to the oscillating X-holes is extremely challenged since these oscillations are highly system dependent and they cannot be anticipated and corrected at the correlation factor level.

Tab. II-5. Atomization energies in kcal/mol for a set of single and multi-bonded molecules. MAE and ME are the mean absolute error and mean error, respectively. For $f_C^{\text{CFX}} f_{\text{XS}}^{5\text{OS}}$, $a = 0.13$ and $\lambda = 1.30$. For $f_C^3 f_{\text{XS}}^{5\text{OS}}$, $\lambda = 0.82$

Mol.	$f_C^{\text{CFX}} f_{\text{XS}}^{5\text{OS}}$	$f_C^{\text{CFX}} f_{\text{XS}}^{4\text{OS}}$	$f_C^3 f_{\text{XS}}^{5\text{OS}}$	$f_C^3 f_{\text{XS}}^{4\text{OS}}$	$f_C^{\text{CFX}} f_X^4$	$f_C^3 f_X^4$	CFX	PBE	exp.
H ₂	109	108	111	111	106	109	108	105	109
LiH	53	54	58	53	51	53	60	53	58
CH ₄	419	438	419	414	422	396	417	419	419
NH ₃	297	321	303	295	293	266	293	300	297
H ₂ O	223	239	231	225	211	196	225	233	232
HF	132	141	138	134	121	115	137	141	141
Li ₂	9	10	15	13	9	13	24	20	24
LiF	124	135	134	127	107	107	141	139	139
Be ₂	6	18	5	1	15	-8	5	10	3
MAE	6.3	9.9	3.1	6.0	13.0	19.4	2.1	2.7	0
ME	-5.5	4.5	-1.0	-5.6	-9.6	-19.4	-0.8	-0.3	0
C ₂ H ₂	410	453	408	397	415	356	406	412	405
C ₂ H ₄	573	626	569	556	591	518	561	570	563
HCN	322	369	323	312	321	259	314	324	312
CO	264	304	267	257	259	208	258	268	259
N ₂	244	297	246	234	242	174	229	241	229
NO	152	205	152	140	162	94	154	170	153
O ₂	102	152	95	86	126	57	126	142	121
F ₂	19	55	36	23	12	-22	35	51	39
P ₂	88	106	116	103	84	84	114	119	117
Cl ₂	38	50	51	42	39	33	54	63	58
MAE	13.4	39.9	8.4	11.5	15.2	49.6	2.2	10.5	0
ME	-4.5	36.2	0.7	-10.6	-0.5	-49.6	-0.5	10.5	0
Tot. MAE	10.0	25.6	5.9	8.9	14.2	35.3	2.1	6.8	0
Tot. ME	-4.9	21.1	-0.1	-8.2	-4.8	-35.3	-0.7	5.4	0

7.3. Static correlation from MCSCF

We also employ the f_X^4 ansatz to reproduce the energy density of MCSCF, resulting in $f_{\text{XS}}^{4\text{MC}}$. The results obtained with this approach are shown in Tables II-6 and II-7. For atomic energies, just as for the static correlation from the OS approach, the results obtained with the f_C^{CFX} correlation factor are better than those obtained with f_C^3 . The latter generally underestimates the atomic energy, thus does not add enough dynamic correlation. The atomization energies obtained with f_C^{CFX} are significantly overestimated. The atomization energies obtained f_C^3 are improved compared to f_C^{CFX} , but are still worse than those obtained with the OS approach from the previous section. The XC hole models bases on the MCSCF calculations are analyzed in Fig. 2.4. Just as observed for the OS approach, the core-valence regions are the most problematic, with considerable artificial oscillations. We do not have access to the exact XC-hole, however, since our CFXS model [22] has yields good results for

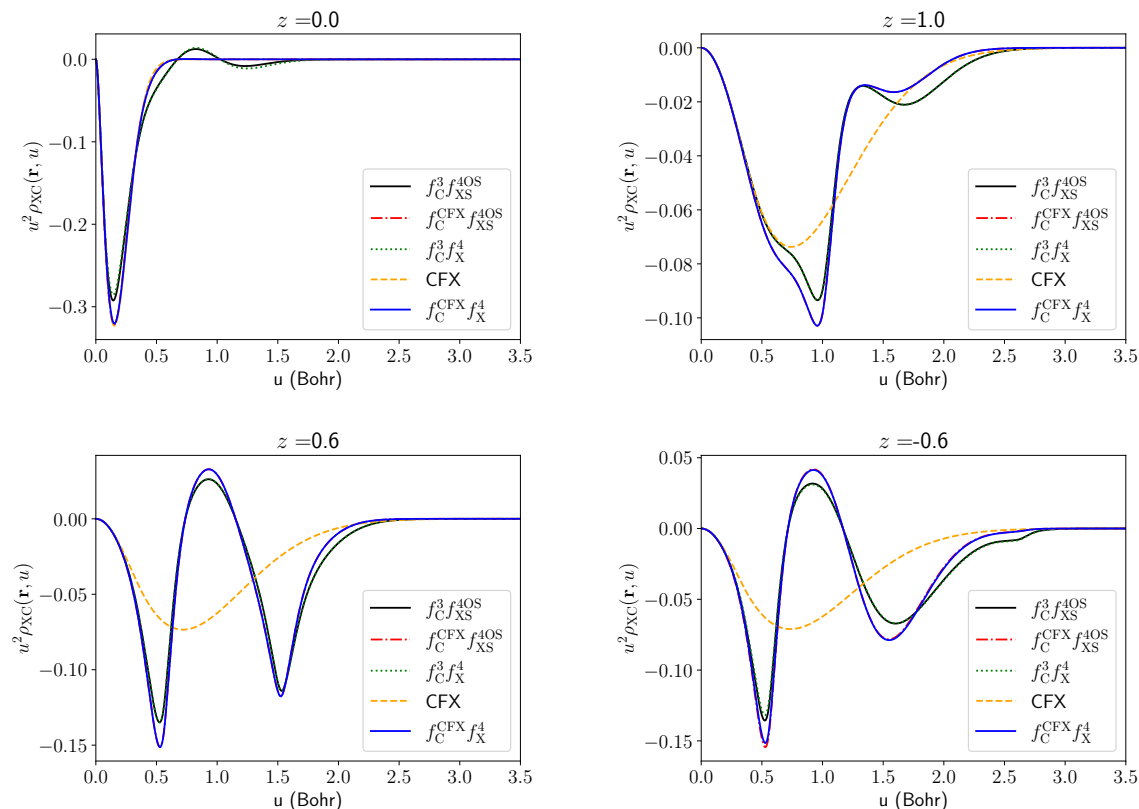


Fig. 2.3. Comparison of XC holes for the N_2 molecule. One N atom is at the position $(0.00, 0.00, 0.00)$ and the other at $(0.00, 0.00, 2.08)$ in Bohr. z is the position along the molecular axis.

various properties, we expect it to provide a good approximation to the XC hole, just like the BR hole is a good representation of the exact X hole [51]. Thus, the main issue for both our models with MCSCF and OS seems to be that a simple exchange factor such as the f_X^4 is not adapted for the complicated oscillatory behavior of $\rho(\mathbf{r}, u)$.

Tab. II-6. Total energies for atoms in Hartree and the Mean Absolute Error(MAE) for different methods. For $f_C^{\text{CFX}} f_{\text{XS}}^{\text{4MC}}$, $a = 0.18$ the same as [21].

Atom	$f_C^{\text{CFX}} f_{\text{XS}}^{\text{4MC}}$	$f_C^3 f_{\text{XS}}^{\text{4MC}}$	CFXS	E_X^{KS}	MCSCF	PBE	exp.
H	-0.500	-0.500	-0.500	-0.499	-0.500	-0.500	-0.500
He	-2.902	-2.904	-2.912	-2.859	-2.860	-2.891	-2.904
Li	-7.467	-7.484	-7.485	-7.431	-7.432	-7.461	-7.478
Be	-14.644	-14.669	-14.664	-14.570	-14.572	-14.628	-14.667
B	-24.642	-24.644	-24.646	-24.528	-24.562	-24.609	-24.654
C	-37.837	-37.803	-37.830	-37.686	-37.704	-37.795	-37.845
N	-54.579	-54.504	-54.567	-54.394	-54.395	-54.530	-54.589
O	-75.083	-74.956	-75.038	-74.802	-74.802	-75.005	-75.067
F	-99.765	-99.588	-99.694	-99.394	-99.396	-99.661	-99.734
Ne	-128.971	-128.748	-128.886	-128.519	-128.527	-128.847	-128.938
Na	-162.313	-162.068	-162.185	-161.839	-161.846	-162.158	-162.255
Mg	-200.103	-199.846	-199.973	-199.600	-199.607	-199.945	-200.053
Al	-242.407	-242.134	-242.227	-241.865	-241.894	-242.228	-242.346
Si	-289.400	-289.116	-289.204	-288.843	-288.861	-289.226	-289.359
P	-341.275	-340.977	-341.062	-340.702	-340.709	-341.106	-341.259
S	-398.109	-397.803	-397.881	-397.491	-397.496	-397.940	-398.110
Cl	-460.131	-459.813	-459.883	-459.467	-459.470	-459.960	-460.148
Ar	-527.511	-527.180	-527.237	-526.800	-526.807	-527.332	-527.540
MAE	0.023	0.151	0.089	0.342	0.334	0.090	-

Tab. II-7. Atomization energies in kcal/mol for a set of single and multi-bonded molecules. MAE and ME are the mean absolute error and mean error, respectively. For $f_C^{\text{CFX}} f_{\text{XS}}^{\text{4MC}}$, $a = 0.18$ the same as [21].

Mol.	$f_C^{\text{CFX}} f_{\text{XS}}^{\text{4MC}}$	$f_C^3 f_{\text{XS}}^{\text{4MC}}$	CFXS	E_X^{KS}	MCSCF	PBE	exp.
H ₂	114	115	112	83	95	105	109
LiH	54	59	61	33	44	53	58
CH ₄	465	434	413	325	370	419	419
NH ₃	343	312	293	197	248	300	297
H ₂ O	250	230	231	152	192	233	232
HF	140	131	143	95	114	141	141
Li ₂	17	22	21	3	10	20	24
LiF	119	112	146	88	94	139	139
Be ₂	43	3	4	-12	-8	10	3
MAE	20.7	8.7	3.3	50.7	29.3	2.7	0
ME	13.8	-0.7	0.2	-50.7	-29.3	-0.3	0
C ₂ H ₂	505	419	401	287	357	412	405
C ₂ H ₄	674	585	556	421	499	570	563
HCN	420	345	311	191	285	324	312
CO	355	288	256	167	251	268	259
N ₂	334	252	229	107	211	241	229
NO	253	174	152	42	128	170	153
O ₂	191	123	121	21	90	142	121
F ₂	70	33	41	-45	16	51	39
P ₂	144	140	124	28	95	119	117
Cl ₂	63	55	56	14	38	63	58
MAE	75.4	17.6	2.7	102.3	28.6	10.5	0
ME	75.4	15.9	-1.0	-102.3	-28.6	10.5	0
Tot. MAE	49.5	13.4	3.0	77.9	28.9	6.8	0
Tot. ME	46.2	8.1	-0.4	-77.9	-28.9	5.4	0

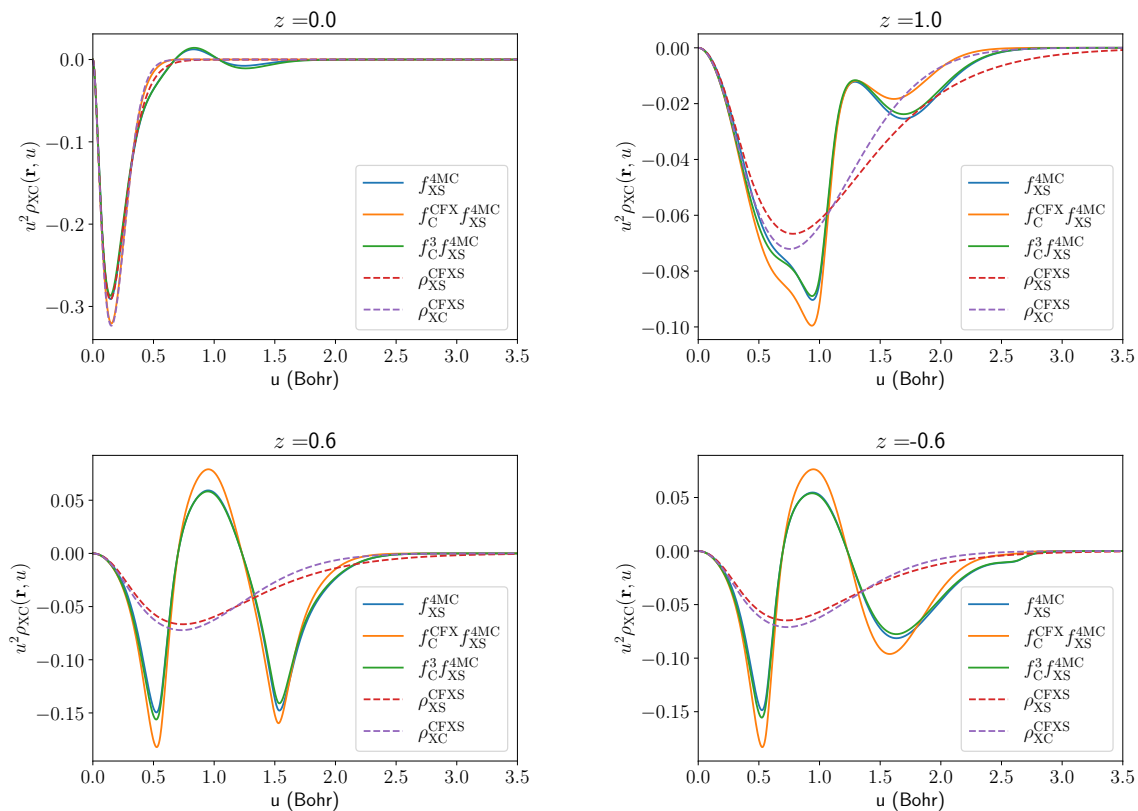


Fig. 2.4. Approximated XC and XS holes for the N₂ molecule. One N atom is at the position (0.00,0.00,0.00) and the other at (0.00,0.00,2.08) in Bohr. z is the position along the molecular axis.

8. Conclusions

Completely non-local approximations to E_{XC} are presently not in general use. There are several reasons for that, one being the increased computational costs in comparison to, for instance, the generalized gradient approximation. While the one-electron, self-interaction error is easy to eliminate within the non-local scheme, the approximations that have been developed up to now are not as accurate as GGAs or hybrids. Here we construct various new exchange and exchange-correlation functionals within the non-local scheme and we analyze their performance. We develop new non-local approximations, building on exchange and correlation factor models previously proposed by our group [18–22]. The non-local model with a single empirical parameter ($f_C^3 f_X^{5OS}$) has a MAE for the calculation of atomization energies of about 6 kcal/mol for a small datasets of molecule. This has to be compared to an MAE of 7 kcal/mol obtained with PBE. To explain the limited performance of the non-local approximations, we perform a detailed analysis of the underlying X and XC holes. We find that $\rho(\mathbf{r}, u)$ exhibits oscillations where the maxima correspond to the core regions of atoms and the minima to mid-bond regions, so that the number of atoms involved increases the oscillations in $\rho(\mathbf{r}, u)$. The X and XC holes on the other hand have much less oscillations and is not obvious how to convert $\rho(\mathbf{r}, u)$ into realistic X and XC holes. Since the exact X hole can be calculated, we can construct the corresponding exchange factor $f_X(\mathbf{r}, u)$ and analyze its behavior. This reveals that the oscillations in $f_X(\mathbf{r}, u)$ compensate for those in $\rho(\mathbf{r}, u)$, it is again not obvious how to model $f_X(\mathbf{r}, u)$.

There are various new elements that we introduced into the non-local approach. We constrain our the approximations to reproduce the exact $\epsilon_X(\mathbf{r})$, or even to reproduce the exchange plus static correlation energy per particle of MCSCF. The constructed exchange and exchange plus static correlation holes are one-electron, self-interaction free and for the cases studied, they are bound from below by $-\rho(\mathbf{r}, u)$. However, the underlying non-locality of $\rho(\mathbf{r}, u)$ is still not sufficiently reduced in our hole models to yield accurate XC energies. Furthermore, we establish two new constraints for $\epsilon_{X,\sigma}$ in terms of $\rho_\sigma(\mathbf{r}, u)$, we provide a lower and an upper bound to $\epsilon_{X,\sigma}$, these bounds coincide for one-electron systems.

9. Data Availability Statement

The data that supports the findings of this study are available within the article.

10. Appendix

10.1. Computation of the on-top value, curvature, and the energy density of the MCSCF XS hole.

Here we detail the procedure employed for the computation of the on-top value and curvature of the MCSCF XS hole and the associated XS energy density, which are ingredients for the models presented in subsection 7.1. We begin by providing a convenient expression for the pair density of an MCSCF wave function $\Psi^{MC} = \sum_I c_I \Phi_I$. After explicitly evaluating the inactive-inactive and inactive-active parts of the 2-particle density matrix $\Gamma_{pqrs} = \sum_{IJ} c_I c_J \langle \Phi_I | \hat{E}_{pq} \hat{E}_{rs} - \delta_{qr} \hat{E}_{ps} | \Phi_J \rangle$, where $\hat{E}_{pq} = \hat{a}_{p\alpha}^\dagger \hat{a}_{q\alpha} + \hat{a}_{p\beta}^\dagger \hat{a}_{q\beta}$, and $\hat{a}_{p\sigma}^\dagger$ and $\hat{a}_{p\sigma}$ are annihilation and creation operators for an electron of spin σ in a spatial molecular orbital (MO) p , the pair density can be expressed as

$$\begin{aligned} P^{MC}(\mathbf{r}, \mathbf{r}') &= \sum_{ij} 4\rho_{ii}(\mathbf{r})\rho_{jj}(\mathbf{r}') - 2\rho_{ij}(\mathbf{r})\rho_{ij}(\mathbf{r}') \\ &+ \sum_{iab} \gamma_{ab} [2\rho_{ii}(\mathbf{r})\rho_{ab}(\mathbf{r}') + 2\rho_{ab}(\mathbf{r})\rho_{ii}(\mathbf{r}') - 2\rho_{ia}(\mathbf{r})\rho_{ib}(\mathbf{r}')] \\ &+ \sum_{abcd} \Gamma_{abcd} \rho_{ab}(\mathbf{r})\rho_{cd}(\mathbf{r}'), \end{aligned} \quad (10.1)$$

with $\gamma_{pq} = \sum_{IJ} c_I c_J \langle \Phi_I | \hat{E}_{pq} | \Phi_J \rangle$ and $\rho_{pq}(\mathbf{r}) = \phi_p(\mathbf{r})\phi_q(\mathbf{r})$ and where indices i, j run over inactive (doubly occupied) orbitals and a, b, c, d run over active orbitals. Using the expression for the pair density of a single closed-shell determinant in terms of its 1-particle reduced density matrix, $P^{SD}(\mathbf{r}, \mathbf{r}') = \gamma(\mathbf{r}, \mathbf{r})\gamma(\mathbf{r}', \mathbf{r}') - \frac{1}{2}\gamma(\mathbf{r}, \mathbf{r}')\gamma(\mathbf{r}', \mathbf{r})$, $P^{MC}(\mathbf{r}, \mathbf{r}')$ can be re-expressed as

$$P^{MC}(\mathbf{r}, \mathbf{r}') = P^{SD}(\mathbf{r}, \mathbf{r}') + \sum_{abcd} \Gamma_{abcd}^S \rho_{ab}(\mathbf{r})\rho_{cd}(\mathbf{r}'), \quad (10.2)$$

where $\Gamma_{abcd}^S = \Gamma_{abcd} - \gamma_{ab}\gamma_{cd} + \frac{1}{2}\gamma_{ac}\gamma_{bd}$. Inserting this expression into the definition of the MCSCF XS hole 4.5 and evaluating the on-top value and curvature from the result yields the readily computable expressions (assuming a small active space such as the ones used in this work)

$$\rho(\mathbf{r})\rho_{XS}^{MC}(\mathbf{r}, u=0) = -\frac{1}{2}\rho(\mathbf{r})\rho(\mathbf{r}) + \sum_{abcd} \Gamma_{abcd}^S \rho_{ab}(\mathbf{r})\rho_{cd}(\mathbf{r}) \quad (10.3)$$

and

$$6\rho(\mathbf{r})\frac{d^2\rho_{XS}(\mathbf{r}, u)}{du^2}\Big|_{u=0} = \frac{1}{2}\rho(\mathbf{r})(\nabla^2\rho(\mathbf{r}) - 2\tau(\mathbf{r}) + \frac{1}{2}|\nabla\rho(\mathbf{r})|^2) + \nabla_{\mathbf{r}'}^2 \sum_{abcd} \Gamma_{abcd}^S \rho_{ab}(\mathbf{r})\rho_{cd}(\mathbf{r}')\Big|_{\mathbf{r}'=\mathbf{r}}, \quad (10.4)$$

with the kinetic energy density $\tau(\mathbf{r}) = \frac{1}{2} \sum_{pq} \gamma_{pq} \nabla \phi_p(\mathbf{r}) \cdot \nabla \phi_q(\mathbf{r})$. Also, from 4.5 and 10.2, the MCSCF XS energy density is given by

$$\begin{aligned} \rho(\mathbf{r})\epsilon_{XS}^{MC}(\mathbf{r}) &= -\frac{1}{4} \int d\mathbf{r}' \frac{\gamma(\mathbf{r},\mathbf{r}')\gamma(\mathbf{r}',\mathbf{r})}{|\mathbf{r}-\mathbf{r}'|} + \frac{1}{2} \sum_{abcd} \Gamma_{abcd}^S \rho_{ab}(\mathbf{r}) \int d\mathbf{r}' \frac{\rho_{cd}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \\ &= \rho(\mathbf{r})(\epsilon_X^{MC}(\mathbf{r}) + \epsilon_S^{MC}(\mathbf{r})). \end{aligned} \quad (10.5)$$

The "exchange" part can be evaluated in the atomic orbitals (AO) basis $\{\chi_\mu\}$ using the expression derived in previous work [20],

$$\rho(\mathbf{r})\epsilon_X^{MC}(\mathbf{r}) = -\frac{1}{4} \sum_{\mu\nu} F_\mu(\mathbf{r})F_\nu(\mathbf{r}) \int d\mathbf{r}' \frac{\chi_\mu(\mathbf{r}')\chi_\nu(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}, \quad (10.6)$$

with $F_\nu = \sum_\mu \gamma_{\mu\nu} \chi_\mu(\mathbf{r})$. The remaining part of the XS energy density can be obtained by transforming $\int d\mathbf{r}' \frac{\chi_\mu(\mathbf{r}')\chi_\nu(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}$ from the AO to the MO basis (only the active subspace is required) and evaluating the following :

$$\rho(\mathbf{r})\epsilon_S^{MC}(\mathbf{r}) = \sum_{abcd} \Gamma_{abcd}^S \phi_a(\mathbf{r})\phi_b(\mathbf{r}) \int d\mathbf{r}' \frac{\phi_c(\mathbf{r}')\phi_d(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}. \quad (10.7)$$

Provided that a small active space is used, the most computationally expensive step for the evaluation of $\epsilon_S^{MC}(\mathbf{r})$ is the transformation of $\int d\mathbf{r}' \frac{\chi_\mu(\mathbf{r}')\chi_\nu(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}$ from the AO basis to the MO basis, which scales as the number of grid points times the number of atomic orbitals squared times the number of active orbitals.

References

- [1] W. Kohn and L. J. Sham, *Physical Review* **140**, A1134 (1965).
- [2] R. G. Parr and W. Yang, Oxford University Press (1989).
- [3] R. M. Dreizler and E. K. Gross, *Density Functional Theory: An Approach to the Quantum Many-Body Problem* (Springer Science & Business Media, 2012).
- [4] A. D. Becke, *Journal of Chemical Physics* **140**, 0 (2014).
- [5] O. Gunnarsson, M. Jonson, and B. I. Lundqvist, *Solid State Communications* **24**, 765 (1977).
- [6] J. A. Alonso and L. A. Girifalco, *Physical Review B* **17**, 3735 (1978).
- [7] O. Gunnarsson, M. Jonson, and B. I. Lundqvist, *Physical Review B* **20**, 3136 (1979).
- [8] Z. Wu, R. E. Cohen, and D. J. Singh, *Physical Review B* **70**, 1 (2004).
- [9] D. J. Singh, *Physical Review B* **48**, 14099 (1993).
- [10] P. García-González, P. García-González, J. E. Alvarellos, E. Chacón, and P. Tarazona, *Physical Review B* **62**, 16063 (2000).
- [11] J. Charlesworth, *Physical Review B* **53**, 12666 (1996).
- [12] M. Sadd and M. P. Teter, *Journal of Molecular Structure: THEOCHEM* **501-502**, 147 (2000).
- [13] K. J. Giesbertz, R. Van Leeuwen, and U. Von Barth, *Physical Review A* **87**, 1 (2013).
- [14] R. Cuevas-Saavedra, D. Chakraborty, S. Rabi, C. Cárdenas, and P. W. Ayers, *Journal of Chemical Theory and Computation* **8**, 4081 (2012).
- [15] R. Cuevas-Saavedra, D. Chakraborty, and P. W. Ayers, *Physical Review A* **85**, 042519 (2012).
- [16] H. Antaya, Y. Zhou, and M. Ernzerhof, *Physical Review A* **90**, 1 (2014).
- [17] T. Tsuneda and K. Hirao, *The Journal of Chemical Physics* **140**, 18A513 (2014).
- [18] J. Přecechtělová, H. Bahmann, M. Kaupp, and M. Ernzerhof, *The Journal of Chemical Physics* **141**, 111102 (2014).
- [19] J. Pavlíková Přecechtělová, H. Bahmann, M. Kaupp, and M. Ernzerhof, *The Journal of Chemical Physics* **143**, 144102 (2015).
- [20] R. Wang, Y. Zhou, and M. Ernzerhof, *The Journal of Chemical Physics* **150**, 084107 (2019).
- [21] R. Wang, Y. Zhou, and M. Ernzerhof, *The Journal of Chemical Physics* **151**, 194102 (2019).
- [22] P.-O. Roy, E. Cuierrier, and M. Ernzerhof, *The Journal of Chemical Physics* **152**, 211101 (2020).
- [23] M. Chan, D. Chakraborty, and P. Ayers, *Computation* **5**, 42 (2017).
- [24] P. Gori-Giorgi, J. G. Ángyán, and A. Savin, *Canadian Journal of Chemistry* **87**, 1444 (2009).

- [25] M. Seidl, P. Gori-Giorgi, and A. Savin, *Phys. Rev. A* **75**, 042511 (2007).
- [26] L. O. Wagner and P. Gori-Giorgi, *Physical Review A - Atomic, Molecular, and Optical Physics* **90**, 1 (2014), arXiv: 1408.5426.
- [27] H. Bahmann, Y. Zhou, and M. Ernzerhof, *The Journal of Chemical Physics* **145**, 124104 (2016).
- [28] S. Vuckovic and P. Gori-Giorgi, *Journal of Physical Chemistry Letters* **8**, 2799 (2017).
- [29] E. Cuierrier, P.-O. Roy, and M. Ernzerhof, *The Journal of Chemical Physics* 10.1063/5.0062940 (forthcoming).
- [30] J. P. Perdew and Y. Wang, *Phys. Rev. B* **46**, 12947 (1992).
- [31] M. Ernzerhof and J. Perdew, *The Journal of Chemical Physics* **109**, 3313 (1998).
- [32] S. Vuckovic, T. J. P. Irons, L. O. Wagner, A. M. Teale, and P. Gori-Giorgi, *Phys. Chem. Chem. Phys.* **19**, 6169 (2017).
- [33] A. D. Becke, *Journal of Chemical Physics* **119**, 2972 (2003).
- [34] A. D. Becke and M. R. Roussel, *Physical Review A* **39**, 3761 (1989).
- [35] J. P. Perdew, A. Savin, and K. Burke, *Physical Review A* **51**, 4531 (1995).
- [36] A. D. Becke, A. Savin, and H. Stoll, *Theoretica Chimica Acta* **91**, 147 (1995).
- [37] J. P. Perdew, M. Ernzerhof, K. Burke, and A. Savin, *International Journal of Quantum Chemistry* **61**, 197 (1997).
- [38] H. Bahmann and M. Ernzerhof, *The Journal of Chemical Physics* **128**, 234104 (2008).
- [39] Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. McClain, E. R. Sayfutyarova, S. Sharma, S. Wouters, and G. K.-L. Chan, arXiv:1701.08223 [physics] (2017).
- [40] Q. Sun, X. Zhang, S. Banerjee, P. Bao, M. Barbry, N. S. Blunt, N. A. Bogdanov, G. H. Booth, J. Chen, Z.-H. Cui, J. J. Eriksen, Y. Gao, S. Guo, J. Hermann, M. R. Hermes, K. Koh, P. Koval, S. Lehtola, Z. Li, J. Liu, N. Mardirossian, J. D. McClain, M. Motta, B. Mussard, H. Q. Pham, A. Pulkin, W. Purwanto, P. J. Robinson, E. Ronca, E. Sayfutyarova, M. Scheurer, H. F. Schurkus, J. E. T. Smith, C. Sun, S.-N. Sun, S. Upadhyay, L. K. Wagner, X. Wang, A. White, J. D. Whitfield, M. J. Williamson, S. Wouters, J. Yang, J. M. Yu, T. Zhu, T. C. Berkelbach, S. Sharma, A. Sokolov, and G. K.-L. Chan, *The Journal of Chemical Physics* **153**, 024109 (2020), arXiv: 2002.12531.
- [41] J. P. Perdew, K. Burke, and M. Ernzerhof, *Phys. Rev. Lett.* **77**, 3865 (1996).
- [42] A. D. Becke, *The Journal of Chemical Physics* **122**, 064101 (2005).
- [43] J. P. Perdew, V. N. Staroverov, J. Tao, and G. E. Scuseria, *Physical Review A* **78**, 052513 (2008).
- [44] P. Mori-Sánchez, A. J. Cohen, and W. Yang, *The Journal of Chemical Physics* **124**, 091102 (2006).
- [45] C. L. Pekeris, *Phys. Rev.* **126**, 1470 (1962).
- [46] C. Schwartz, *International Journal of Modern Physics E* **15**, 877 (2006).

- [47] S. J. Chakravorty, S. R. Gwaltney, E. R. Davidson, F. A. Parpia, and C. F. P. Fischer, *Phys. Rev. A* **47**, 3649 (1993).
- [48] J. A. Pople, M. Head-Gordon, D. J. Fox, K. Raghavachari, and L. A. Curtiss, *The Journal of Chemical Physics* **90**, 5622 (1989).
- [49] L. A. Curtiss, C. Jones, G. W. Trucks, K. Raghavachari, and J. A. Pople, *The Journal of Chemical Physics* **93**, 2537 (1990).
- [50] V. E. Bondybey and J. H. English, *The Journal of Chemical Physics* **80**, 568 (1984).
- [51] Y. Wang, E. Proynov, and J. Kong, *The Journal of Chemical Physics* **154**, 024101 (2021).

Deuxième article.

Constructing and representing exchange–correlation holes through artificial neural networks

par

Etienne Cuierrier¹, Pierre-Olivier Roy¹, Matthias Ernzerhof¹

(¹) Département de Chimie, Université de Montréal, C.P. 6128 Succursale A, Montréal, Québec H3C 3J7, Canada

Cet article a été publié dans *The Journal of Chemical Physics*, 155, 174121 (2021); <https://doi.org/10.1063/5.0062940>.

Les principales contributions de Étienne Cuierrier à cet article sont présentées. J'ai fait la construction des modèles, le code pour entraîner les réseaux de neurones et les calculs. J'ai aussi contribué à l'écriture de l'article et à l'analyse des résultats. Pierre-Olivier Roy a fait le code pour calculer les énergies d'échange des atomes, a contribué à l'écriture de l'article et à l'analyse des résultats.

Cet article concerne la construction de trous d'échange-corrélation à l'aide de réseaux de neurones artificiels. Le trou d'échange-corrélation est une propriété clé en théorie de la fonctionnelle de la densité, il permet de décrire les interactions quantiques entre les électrons. Néanmoins, il n'existe pas de formule mathématique simple pour calculer le trou d'échange-corrélation, ainsi des approximations sont utilisées en pratique pour faire les calculs.

Dans le projet du premier article de cette thèse, il est montré qu'il est difficile de construire des approximations du trou d'échange-corrélation avec la densité électronique non locale et des équations mathématiques simples. Ainsi, pour le projet de ce deuxième article, nous nous sommes intéressés à un outil qui gagne en popularité dans tous les domaines de science: l'apprentissage automatique. Avant de nous attaquer à une propriété complexe comme la densité électronique non locale, pour cet article, nous nous sommes intéressés à développer des approximations avec la densité électronique locale. Nous avons montré qu'à l'aide de contraintes physiques et mathématiques, ainsi qu'un réseau de neurones artificiels, il est possible de construire automatiquement des approximations au trou d'échange-corrélation. Le calcul d'énergie d'échange de quelques atomes montre le potentiel de notre approche pour construire des trous d'échange-corrélation.

De plus, dans cet article, nous avons souhaité à améliorer les résultats du facteur de corrélation qui est développé par le groupe. Ce facteur est multiplié par une approximation du trou d'échange et ce trou d'échange ne respecte pas toutes les contraintes que nous aimerions qu'il satisfasse. Ainsi, dans ce projet, nous nous sommes servis d'un réseau de neurones afin d'ajouter une contrainte à un trou d'échange existant, tout en respectant les contraintes que ce trou respecte déjà. Cette étude montre donc tout le potentiel de l'apprentissage automatique pour construire et améliorer des approximations du trou d'échange-corrélation.

ABSTRACT. One strategy to construct approximations to the exchange-correlation energy E_{XC} of Kohn-Sham density functional theory relies on physical constraints satisfied by the exchange-correlation (XC) hole $\rho_{XC}(\mathbf{r}, u)$. In the XC hole, the reference charge is located at \mathbf{r} and u is the electron-electron separation. With mathematical intuition, a given set of physical constraints can be expressed in a formula, yielding an approximation to $\rho_{XC}(\mathbf{r}, u)$ and the corresponding E_{XC} . Here we adapt machine learning algorithms to partially automate the construction of X and XC holes. While machine learning usually relies on finding patterns in data sets, does not require physical insight, we focus entirely on the latter and develop a tool (ExMachina), consisting of the basic equations and their implementation, for the machine generation of approximations. To illustrate ExMachina, we apply it to calculate various model holes and show how to go beyond existing approximations.

Keywords: Kohn-Sham Density Functional Theory, Machine Learning, Exchange-Correlation Energy

1. Introduction

Machine learning (ML) is rapidly gaining popularity in various fields of science. In density functional theory (DFT) [1–4], methods using ML are being developed at a fast pace (for recent reviews we refer to [5–7]). Since the solution of the Kohn-Sham (KS) equation [1] is usually the computationally limiting step in applications of DFT, ML models are developed to bypass it [8–11]. ML is also applied to improve the accuracy of approximations to the exchange-correlation (XC) energy E_{XC} [12–18]. Nagai and co-workers [13] showed that semi-local approximations to E_{XC} can be accurately represented with multi-layer, artificial neural networks (NNs). They developed improvements to semi-local functionals, taking advantage of automatic differentiation in the PyTorch software library [19] to obtain the KS or generalized KS potentials. Self-consistent implementations of functionals are thus easy to accomplish with NNs.

In general, in DFT, ML is employed as a new tool for fitting approximate functionals to accurate experimental or theoretical data. However, in wave function theory (WFT), NNs are also used to obtain solutions to the Schrödinger equation [6, 20–22]. Representing wave-functions through NNs is complicated by the fact that usually a large number of solutions are obtained, each corresponding to a local minimum of a loss function [23].

In the present work, we also use NNs to tackle involved physical equations. These equations, which we develop within this work, have approximate XC holes as solutions. These solutions might not be unique but are physically acceptable and useful. In the remainder of the introduction, we discuss selected unsolved problems of functional development that can be addressed with ML.

There are various strategies for the development of improved approximations to E_{XC} , at the extreme opposing ends are first-principles and empirical approaches. In the former, exact conditions and equations are employed to guide the construction of formulas and procedures and, in the latter, data are fitted through appropriate formulas. The research pursued in our group focuses on the constraints-based approximations to the XC hole. Often, we employ the factorization ansatz [24–29] where, for example, the XC hole

$$\rho_{XC}(\mathbf{r},u) = f(\mathbf{r},u)\rho_X(\mathbf{r},u) \tag{1.1}$$

is a product of the correlation factor $f(\mathbf{r},u)$ and the X hole model $\rho_X(\mathbf{r},u)$. The reference point is \mathbf{r} and u denotes the electron-electron distance. In actual approximations, both of these factors $f(\mathbf{r},u)$ and $\rho_X(\mathbf{r},u)$ have to be modeled and here we provide a ML tool for these tasks. Similarly, there are various other methods to construct approximate E_{XC} that rely on models for $\rho_X(\mathbf{r},u)$ and $\rho_{XC}(\mathbf{r},u)$, such as range-separation [30–38]. In these contexts, a key problem is to derive a flexible model for $\rho_X(\mathbf{r},u)$ that can reproduce a given exchange energy per particle $\epsilon_X(\mathbf{r})$. More generally, the conditions that we impose on $\rho_{XC}(\mathbf{r},u)$, $f(\mathbf{r},u)$,

and $\rho_X(\mathbf{r}, u)$ are increasingly numerous and it is laborious, often impossible, for humans to find an explicit formula that satisfies them. ExMachina, the tool developed here, removes this roadblock. It can transform constraints into new XC holes and it can add additional constraints to existing XC holes.

The self-consistent implementation of complicated functionals is another very cumbersome step that limits progress in the development of improved versions of E_{XC} . ExMachina eliminates this problem since the calculations of the needed derivatives are an integral part of machine learning software packages such as PyTorch [19].

2. Theory for generating exchange hole models from physical constraints

To develop the theory supporting the partial automation of approximating $\rho_{XC}(\mathbf{r}, u)$, first we focus on the spherically-averaged X hole $\rho_X(\mathbf{r}, u)$ [39], providing an expression to the exchange energy per particle,

$$\epsilon_X(\mathbf{r}) = \int d^3u \frac{\rho_X(\mathbf{r}, u)}{2u}, \quad (2.1)$$

which integrates to $E_X = \int d^3r \rho(\mathbf{r}) \epsilon_X(\mathbf{r})$. For most systems, the X hole represents a complicated function and various known physical constraints provide partial information about it, albeit insufficient for a unique characterization. Typically this problem is circumvented by employing or adapting the X hole of a particular system, such as the homogeneous electron gas [1] or the hydrogen atom [40]. Here, we avoid biasing our models towards a particular system, thus we determine an entire function from just a few constraints. We make virtue of necessity and find the X hole by maximizing its “unknown-ness” subject to a finite number of constraints. To this end, we define the information entropy S of the X hole,

$$S[\rho_X] = \int_0^\infty du \rho_X(\mathbf{r}, u) \log[-\rho_X(\mathbf{r}, u)]. \quad (2.2)$$

The constraints considered here are of the general form, $\mathfrak{C}^{(i)}[\rho_X] - \mathfrak{c}^{(i)} = 0$, various examples are provided in Sec. 3. $\mathfrak{C}^{(i)}[\rho_X]$ represent a suitable functional of ρ_X , yielding a property of ρ_X which is constrained to have the value $\mathfrak{c}^{(i)}$. These expressions are then combined with $S[\rho_X]$ through the Lagrange multiplier method,

$$L[\rho_X] = -S[\rho_X] + \sum_i^N \lambda_i (\mathfrak{C}^{(i)}[\rho_X] - \mathfrak{c}^{(i)}) \quad (2.3)$$

and, to maximize the constrained entropy (unknown-ness), stationarity is imposed on the functional $L[\rho_X]$,

$$\frac{\delta L[\rho_X]}{\delta \rho_X} = 0, \quad \frac{dL[\rho_X]}{d\lambda_i} = 0, \quad i = 1, 2, \dots, N. \quad (2.4)$$

The equations 2.3 and 2.4 represent a formal basis for the automated construction of X hole models. The functional $S[\rho_X]$ in Eq. 2.2 can be thought of as a weighted differential entropy [41], weighted because the differential volume element of the spherical coordinates ($u^2 du$) has been multiplied by u^{-2} , as further discussed in Sec. 3.1. The constraints represent the physical content of the X hole model, whereas $S[\rho_X]$ is one of various conceivable options to ensure a smooth representation of the constraints through a NN. As the number of constraints imposed grows, the X hole model is increasingly determined through them and the artificial impact of $S[\rho_X]$ is reduced. Traditionally, X and XC hole models are based on the researcher’s intuition and ability to express physical insight in tractable formulas; here, instead, we solve the stationary problem Eq. 2.4 with a machine learning technology, which can be described as label-free supervised approach [42]. We train our NN to reproduce physical properties of the exact X hole, instead of minimizing the error compared to a dataset of X hole values.

The formal solution to Eq. 2.4 is

$$\rho_X(u, \mathbf{c}^{(1)}, \mathbf{c}^{(2)}, \dots, \mathbf{c}^{(N)}) = -e^{-1} \times e^{-\sum_i^N \lambda_i \frac{\delta \mathbf{c}^{(i)}[\rho_X]}{\delta \rho_X}}. \quad (2.5)$$

At first sight, this appears to provide a straightforward expression for the X hole, however, in practice there are numerous problems with Eq. 2.5, exemplified in Sections 3.1 and 3.2. In general, each λ_i depends on all $\{\mathbf{c}^{(i)}\}_{i=1, \dots, N}$ and finding an efficient representation for $\rho_X(u, \mathbf{c}^{(1)}, \mathbf{c}^{(2)}, \dots, \mathbf{c}^{(N)})$ is an exercise that is usually limited to very few constraints or variables $\mathbf{c}^{(i)}$. With the NN-based ExMachina tool, we succeed in generating X hole models as a smooth analytical expression of several variables $\mathbf{c}^{(i)}$. Thus, we overcome a limitation that we previously encountered in applications of the maximum entropy principle to the construction of X hole models [39]. This work, in turn, was drawing on earlier applications of the maximum entropy principle to reconstruct intracule densities [43].

3. Examples of X hole models represented through neural networks

ExMachina relies on physical constraints and therefore, we compile those [39] that are relevant for the hole models in this work:

$$\rho_X(\mathbf{r}, u) \leq 0, \quad (\text{Condition 0})$$

$$\rho_X(\mathbf{r}, u = 0) = -\frac{1}{2}\rho(\mathbf{r}), \quad (\text{Condition I})$$

$$\left. \frac{d\rho_X(\mathbf{r}, u)}{du} \right|_{u=0} = 0, \quad (\text{Condition II})$$

$$\left. \frac{d^2 \rho_X(\mathbf{r}, u)}{du^2} \right|_{u=0} = \frac{1}{6} \left(-\nabla^2 \rho(\mathbf{r}) + 4 \left[\tau(\mathbf{r}) - \frac{|\nabla \rho(\mathbf{r})|^2}{8\rho(\mathbf{r})} \right] \right), \quad (\text{Condition III})$$

$$4\pi \int duu^2 \rho_X(\mathbf{r},u) = -1, \quad (\text{Condition IV})$$

$$4\pi \int duu^2 \frac{\rho_X(\mathbf{r},u)}{2u} = \epsilon_X(\mathbf{r}), \quad (\text{Condition V})$$

$$\left. \frac{d^{2n+1} \rho_X(\mathbf{r},u)}{du^{2n+1}} \right|_{u=0} = 0, \quad n = 0,1,2,\dots \quad (\text{Condition VI})$$

In condition (C) III, in addition to the electron density $\rho(\mathbf{r})$, the curvature of the X hole depends on the positive semi-definite kinetic energy density $\tau(\mathbf{r}) = \sum_i^{occ} (\nabla \psi_i(\mathbf{r}))^2$, where ψ_i is a KS spatial orbital. Condition V involves a given energy density per particle $\epsilon_X(\mathbf{r})$ that is to be reproduced by the X hole. Throughout the article, we consider spin-unpolarized densities. To arrive at the spin-density versions of our approximations, the usual spin-scaling relations [39] are applied, $\rho_X[\rho_\uparrow, \rho_\downarrow](\mathbf{r}, u) = \sum_s \frac{\rho_s(\mathbf{r})}{\rho(\mathbf{r})} \rho_X[2\rho_s](\mathbf{r}, u)$ where s is spin \uparrow or \downarrow . Various of the above listed relations are considered in the widely used X hole models (e.g. [39, 40]). In addition to explicit conditions, through the choice of analytic neural networks, we limit the space of $\rho_X(\mathbf{r}, u)$ to analytic functions. This is an assumption that is often evoked for X hole models and then, since the spherically averaged hole is considered, all odd powers vanish in the Taylor series expansion of $\rho_X(\mathbf{r}, u)$. This means that C-VI is satisfied. C-VI encompasses C-II and we don't need to enforce C-II if C-VI is fulfilled. Here we limit our models to satisfy C-II and ignore C-VI. If the cusps at the nuclei are correctly reproduced, which is not the case in Gaussian basis set codes, then the X hole is non-analytic and C-II and C-VI are in general not satisfied. The popular Becke-Roussel (BR) X hole [40] is non-analytic, it is the exact hole of a hydrogenic atom.

For simple X hole models satisfying only the most basic conditions, the constrained maximization of the entropy can be performed analytically and first we discuss such an example before turning to NNs to represent the X hole models.

3.1. Simplest X hole model employing the on-top and normalization condition

We consider C-I and C-IV and find the explicit solution to the minimization of the $L[\rho_X]$ functional,

$$\rho_X(u, \mathbf{c}^{(1)}, \mathbf{c}^{(4)}) = -e^{-1} e^{-\lambda_1 \delta(u)} e^{-\lambda_4 u^2}, \quad (3.1)$$

where $\mathbf{c}^{(1)} = -\rho/2$ and $\mathbf{c}^{(4)} = -1$. To satisfy the on-top condition C-1, we require that $-e^{-1} e^{-\lambda_1 \delta(u=0)} = -\rho/2$; this is a local condition that has no significance other than at $u = 0$. It also illustrates the problem that, in general, we have to chose a u -range over which we want the condition to impact the X hole. In the example at hand this can be done by replacing the δ function in the exponent by a normalized Gaussian of finite but large width, so that the Gaussian doesn't vary over the range of the X hole and can be replaced by the constant

C , i.e., $\rho_X(u, \mathbf{c}^{(1)}, \mathbf{c}^{(4)}) = -e^{-1} e^{-\lambda_1 C} e^{-\lambda_4 u^2}$. λ_1 is then determined to satisfy C-0, yielding,

$$\rho_X(u, -\rho/2, \mathbf{c}^{(4)}) = -\frac{\rho}{2} e^{-\lambda_4 u^2}. \quad (3.2)$$

Subsequently, λ_4 can be readily calculated by imposing C-4, leading to

$$\rho_X(u, -\rho/2, -1) = -\frac{1}{2} \rho e^{-\pi(2/\rho)^{-2/3} u^2}. \quad (3.3)$$

The Gaussian function is frequently used as a simple approximation to the X hole, but this has been done without realizing that it is the least biased choice in the sense of maximum entropy. As already mentioned, the entropy of the X hole is calculated considering $\rho_X(\mathbf{r}, u)$ to be a one-dimensional function

$$S[\rho_X] = \int_0^\infty du \rho_X(\mathbf{r}, u) \log[-\rho_X(\mathbf{r}, u)], \quad (3.4)$$

and therefore omitting the u^2 factor that would appear in 3 dimensional volume element. Omitting this factor assigns an equal weight over the entire u range, whereas the inclusion of u^2 suppresses the small- u contributions to the entropy. The small- u region is, however, the one where most of the conditions have to be combined to yield a smooth function.

3.2. ExMachina: Neural-network-generated X holes depending on the curvature

Now we focus on the more challenging implementation of C-0 to C-IV. Formally, through Eq. 2.5, the X hole is

$$\rho_X(u, \mathbf{c}^{(1)}, \mathbf{c}^{(2)}, \mathbf{c}^{(3)}, \mathbf{c}^{(4)}) = -e^{-1} \times e^{-\lambda_1 \delta(u=0)} e^{-\lambda_2 \frac{d}{du} \delta(u=0)} e^{-\lambda_3 \frac{d^2}{du^2} \delta(u=0)} e^{-\lambda_4 u^2}. \quad (3.5)$$

The Dirac- δ functions in this expression show $\rho_X(u, \mathbf{c}^{(1)}, \mathbf{c}^{(2)}, \mathbf{c}^{(3)}, \mathbf{c}^{(4)})$ is highly non-analytic and of little use. In the NN approach, by design, we search for analytic solutions for $\rho_X(u, \mathbf{c}^{(1)}, \mathbf{c}^{(2)}, \mathbf{c}^{(3)}, \mathbf{c}^{(4)})$ and this implicit condition is similar to replacing the δ functions by Gaussians of certain widths, as it was done in the previous subsection. In the present context, we interpret an NN as a means to represent a complicated function. It depends on a number of input variables, such as, the electron-electron separation u and the curvature of the X hole, and its function value is the corresponding X hole. To train the NN (optimize its internal parameters), we need to introduce a loss function whose minimization defines the particular X hole model. The steps that result in an X hole model are thus to (I) identify the constraints to be satisfied, (II) state the variables of the X hole model or equivalently of the NN, (III) define the architecture of the NN used to represent the X hole, and lastly (IV), we have to design the loss function whose minimization determines the parameters of the NN. Following these steps, now we generate an approximation that satisfies conditions C-0 through C-IV. Density scaling relations [39] enable us to introduce a reduced X hole $J_X(y)$

that does not explicitly depend on the electron density,

$$\rho_X(u) = \rho J_X(y), \text{ where } y = k_F u \text{ and } k_F = (3\pi^2 \rho)^{1/3}. \quad (3.6)$$

The conditions listed above can be straightforwardly converted into conditions for $J_X(y)$ and the actual computations are carried out with the reduced hole.

In $\rho_X(u, \mathbf{c}^{(1)}, \mathbf{c}^{(2)}, \mathbf{c}^{(3)}, \mathbf{c}^{(4)})$ the values of $\mathbf{c}^{(1)}, \mathbf{c}^{(2)}$, and $\mathbf{c}^{(4)}$ are kept fixed and $\mathbf{c}^{(3)}$ (curvature at the origin), whose index we drop, is the actual variable,

$$\rho_X^{\text{ExMc}}(u, \mathbf{c}) = \rho_X(u, \mathbf{c}^{(1)} = -\rho/2, \mathbf{c}^{(2)} = 0, \mathbf{c}^{(3)} = \mathbf{c}, \mathbf{c}^{(4)} = -1). \quad (3.7)$$

Having specified the input variables (u and \mathbf{c} , or equivalently the reduced variables $y = k_F u$ and $c = d^2 J_X(y = 0)/dy^2 = (\rho k_F^2)^{-1} \mathbf{c}$), as well as the output value of the NN, i.e., the reduced X hole, next we define the architecture of the NN representing $\rho_X^{\text{ExMc}}(u, \mathbf{c})$. As shown in Fig. 3.1, the first layer consists of two neurons representing y and c . The next three hidden layers have 100 neurons each and the final output layer consists of one neuron. The softplus activation function ($\text{Softplus}(x) = \log(1 + e^x)$) are applied to all but the first layer. The parameters of the NN are the standard weights and biases [23] and they are described further in Section 4.1.

Now we construct the loss function employed in the training (determination of weights and biases) of the NN. The X hole that we want to represent with an NN is defined as stationary point of the functional

$$L[\rho_X(\mathbf{c})] = -S[\rho_X(\mathbf{c})] + \sum_{i=1}^4 \lambda_i (\mathbf{e}^{(i)}[\rho_X(\mathbf{c})] - \mathbf{c}^{(i)}). \quad (3.8)$$

While the u variable is integrated over in the functional $L[\rho_X(\mathbf{c})]$, the \mathbf{c} dependence remains and is made evident. Furthermore, $L[\rho_X(\mathbf{c})]$ depends on the curvature $\mathbf{c} = \mathbf{c}^{(3)}$ through the constraint C-III. The loss function $L^p[\rho_X]$ is obtained by implementing $L[\rho_X(\mathbf{c})]$ for a range of \mathbf{c} values through the penalty method [44],

$$L^p[\rho_X] = \sum_{j=1}^N \left(-S[\rho_X(\mathbf{c}_j)] + \sigma \left((\mathbf{e}^{(3)}[\rho_X(\mathbf{c}_j)] - \mathbf{c}_j)^2 + \sum_{i=1,2,4} (\mathbf{e}^{(i)}[\rho_X(\mathbf{c}_j)] - \mathbf{c}^{(i)})^2 \right) \right). \quad (3.9)$$

On the right-hand side of this equation, there is a summation over a discretized set of N curvature points, which implies that the NN is trained for a range of curvature values and it thus represents a function of the curvature. We use a curvature (c) grid of 40 points uniformly distributed over an interval of $[-16, 4]$ in reduced units. This range was chosen as it contains more than 99% of the curvature values for the set of atoms in Table III-1. C-0 does not appear as a constraint in Eq. 3.9, since it is satisfied by multiplying the strictly positive activation function softplus [23], used for the output layer, with -1 .

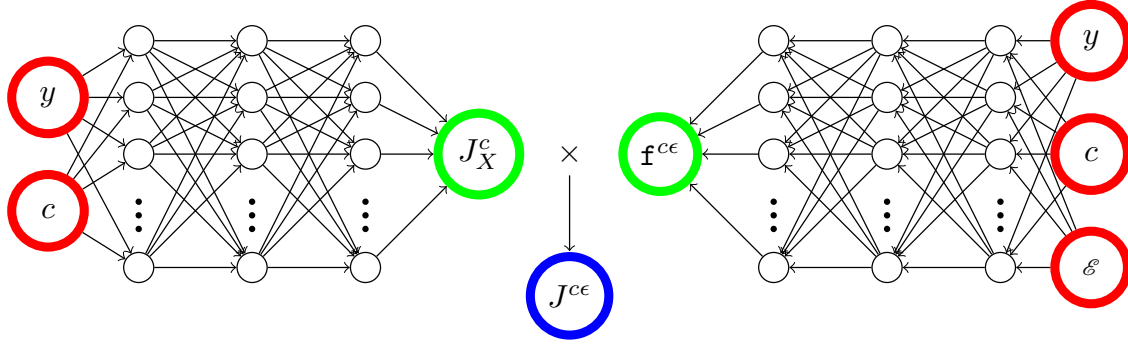


Fig. 3.1. Neural networks (NNs) for the representation of X and XC hole models. The NN on the left represents a reduced X hole $J_X^c(y,c)$ that depends on the reduced curvature (c) and electron separation (y). To satisfy further conditions, including conditions that introduce correlation effects, a second NN representing a correlation and/or exchange factor $\mathbf{f}^{c\epsilon}(y,c,\epsilon)$ is introduced. It transforms $J_X^c(y,c)$ into a new reduced hole model $J^{c\epsilon}(y,c,\epsilon) = \mathbf{f}^{c\epsilon}(y,c,\epsilon) \times J_X^c(y,c)$. In this example, the factor depends on a reduced energy density per particle $\epsilon = \epsilon/|\epsilon_X^{\text{LDA}}|$, ($\epsilon_X^{\text{LDA}} = -\frac{3}{4} \left(\frac{3}{\pi}\right)^{4/3} \rho^{1/3}$). $J_X^c(y,c)$ does not need to be represented by a NN, it can also be a conventional hole model. The final X or XC hole model is then $\rho_{X(C)}(u,\mathbf{c},\epsilon) = \rho \times J^{c\epsilon}(k_F u, \frac{c}{\rho k_F^2}, \frac{\epsilon}{|\epsilon_X^{\text{LDA}}|})$, where $k_F = (3\pi^2 \rho)^{1/3}$. Further details about the NNs are provided in Section 4.1.

During the training cycles, σ in Eq. 3.9 is increased linearly at each optimization step until the mean squared error (MSE) is lower than 10^{-5} for each constraints in reduced units. The integrations over $y = k_F u$ are performed using the Gauss-Legendre scheme with 300 integration points over an interval $[0,25]$ in y . The discretized grid points of the y and c variables represent the data employed in training of the NN, this data is generated from a mathematical expression (Eqs. 2.3 and 2.4) rather than employing empirical data. We use the PyTorch [19, 23] software for all our NNs and their training. Furthermore, the python code for ExMachina is made available on Gitlab at <https://gitlab.com/electronic-structure-udem/exmachina>.

In Fig. 3.2, we compare $\rho_X^{\text{ExMc}}(u,\mathbf{c})$ and the Becke-Roussel $\rho_X^{\text{BR}}(u,\mathbf{c})$ [40]. To evaluate $\rho_X^{\text{BR}}(u,\mathbf{c})$ we employ its parametrization of reference [45]. Here and in the following, we discuss the results in terms of the reduced variables y , c , and the reduced holes J_X of Eq. 3.6. For the curvature values shown, the BR X hole depends much more strongly on this variable than ExMc. Furthermore, whereas $J_X^{\text{BR}}(y,c)$ is highly unsymmetrical as a function of c (negative c -values significantly deepen the hole, while positive c -values have little opposite effect), this is much less so for $J_X^{\text{ExMc}}(y,c)$. We attribute these differences to the fact that the BR hole is not only exploiting the curvature, but it is the X hole of the hydrogen atom and thus $J_X^{\text{BR}}(y,c)$ is completely determined by the curvature, which in this model can be interpreted as a measure for the separation of the reference point from the nucleus. In ExMc, there is no nucleus present, neither implicitly nor explicitly. In our NN implementation, the

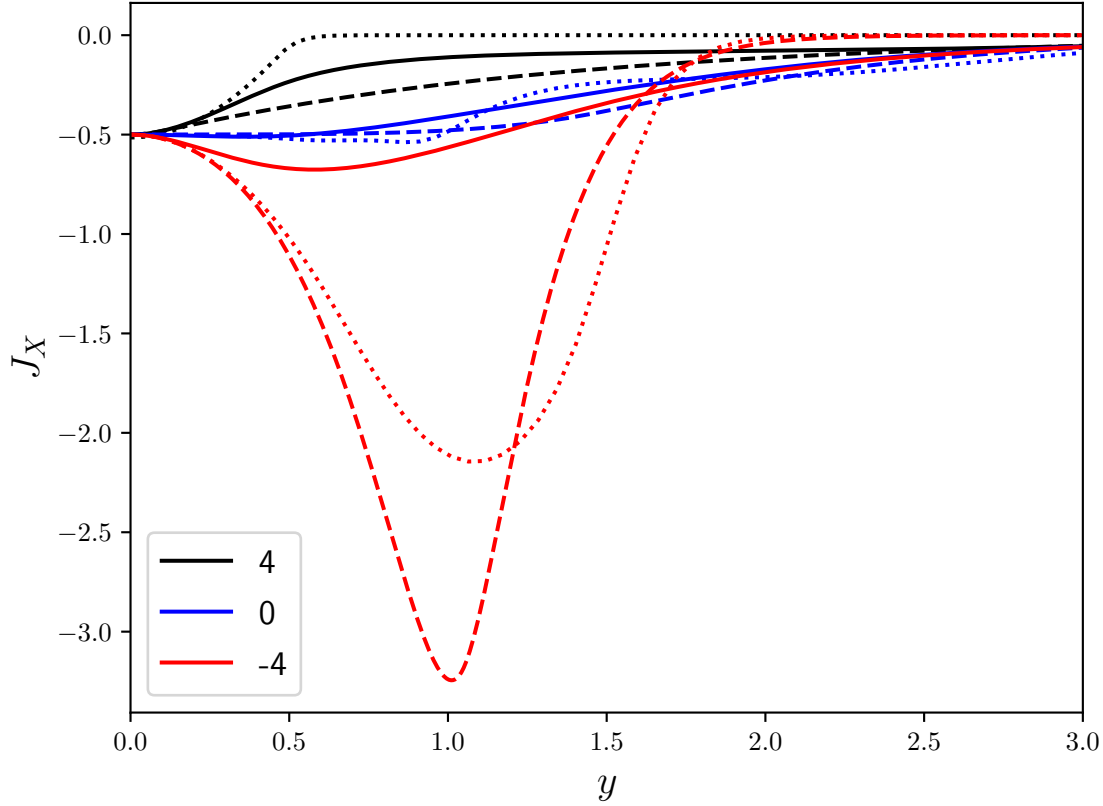


Fig. 3.2. Comparison of $J_X^{\text{ExMc}}(y, c)$ (full lines), $J_X^{\text{ExMc}\alpha}(y, c)$ (dotted lines), and the Becke-Roussel $J_X^{\text{BR}}(y, c)$ (dashed lines) for various values of the reduced curvature $c = -4, 0, 4$. The reduced holes are plotted where $\rho_X(u, \mathbf{c}) = \rho_{J_X}(k_F u, \frac{c}{\rho k_F^2})$.

curvature condition is applied at a single point ($y = 0$) and the analyticity and the details of the NN implicitly determine the range over which C-III is relevant. To explicitly determine the range over which C-III is imposed, we can use the modified Lagrange multiplier term

$$\lambda_3 \int du e^{-\alpha u^2} \left(\frac{d^2}{du^2} \rho_X(\mathbf{r}, u, \mathbf{c}^{(3)}) - \mathbf{c}^{(3)} \right), \quad (3.10)$$

in which $\alpha^{1/2}$ determines the length scale for C-III. Here, our aim is to introduce an NN driven approach to the first-principles construction of approximations rather than perfecting a particular approximation; so, for the purpose of illustration, we set $\alpha = 60$ and re-train the NN to obtain an exchange functional $\text{ExMc}\alpha$. The Lagrange multiplier term Eq. 3.10 is also converted into a penalty term, $\sigma \int du e^{-\alpha u^2} \left(\frac{d^2}{du^2} \rho_X(\mathbf{r}, u, \mathbf{c}^{(3)}) - \mathbf{c}^{(3)} \right)^2$ and included into the loss function Eq. 3.9.

As shown in Fig. 3.2, the resulting $J_X^{\text{ExMc}\alpha}(u, c)$ is closer to $J_X^{\text{BR}}(u, c)$ for $c < 0$. To further analyze our results, in Fig. 3.3, we compare the reduced X energy $\varepsilon_X = \epsilon_X / |\epsilon_X^{\text{LDA}}|$, (

$\epsilon_X^{\text{LDA}} = -\frac{3}{4} \left(\frac{3}{\pi} \right)^{4/3} \rho^{1/3}$) of ExMc, ExMc α , and BR as well as PBEc. LDA denotes the local density approximation and PBEc is obtained by replacing the reduced density gradient in the PBE [46] by the curvature using the relation 27 of Ref. [39] for $c \leq 0.2$. For $c > 0.2$, where the PBE is undefined, we simply employ its $c = 0.2$ value. Such a continuation is needed to ensure that we can perform calculations of X energies with the c -dependent PBEc. The PBEc and BR ϵ_X are comparable to each other whereas ExMc is quite different from the former. The dependence of ExMc on c is weaker than for BR and PBEc. In particular, the reference functionals vary rapidly for $c \approx 0$. ExMc α , however, is surprisingly close to BR for c around zero but more anti-symmetric with respect to the point defined by ϵ_X of ≈ -1 and $c = 0$. Interestingly, the various curves all approximately pass through this point. Moreover, ϵ_X^{BR} shows a rapid saturation as c increases. This is because the curvature of $\rho(\mathbf{r})$ quickly grows towards infinity upon approaching the nucleus of the hydrogenic atom without much change in the corresponding $J_X^{\text{BR}}(u,c)$ and ϵ_X^{BR} . The saturation in ϵ_X^{PBEc} as c increases is a consequence of our choice of continuation of ϵ_X^{PBEc} to $c > 0.2$.

Also shown in Fig. 3.3 is the entropy of the reduced X holes; for negative c , as expected, the entropy of ExMc is higher than that of the other approximations. The PBEc curve for positive values of c is a continuation as described above, thus the entropy of ExMc does not drop below that of PBEc but below that of PBEc at $c = 0.2$ as c increases. The entropy of $J_X^{\text{BR}}(u,c)$ saturates quickly for $c \geq 0$. The BR X hole, being derived from an electron density with cusp, is not analytic, in contrast to our NN models. Satisfying the implicit condition of analyticity by ExMc and ExMc α lowers the resulting entropy compared to BR. Similarly, ExMc α is more constrained than ExMc and this is also visible in its entropy.

As an advantage of the NN representation, self-consistency is easy to implement since the PyTorch system provides the required derivatives. Here we employ the generalized KS scheme, where the energy is made stationary with respect to the occupied orbitals, and we implement it using adapted modules from the PySCF package [47, 48]. All reported calculations are performed self-consistently, as described, with cc-pvtz basis sets. To do so, the reduced exchange energy densities per particle of ExMc, ExMc α , and BR are fitted as a function of the curvature of the reduced exchange hole using a neural network with 3 hidden layers (100 neurons each) with exponential linear unit (ELU) [23] activation functions. The functionals are represented on a grid of reduced curvatures ranging from -16 to 4, using approximately 2000 uniformly distributed points. The NNs are optimized by minimizing the mean square relative error (MSRE) until it reached values less than 1×10^{-7} . The local density approximation results are obtained conventionally. The exact exchange results are obtained from self-consistent Hartree-Fock calculations, consistent with the generalized KS scheme.

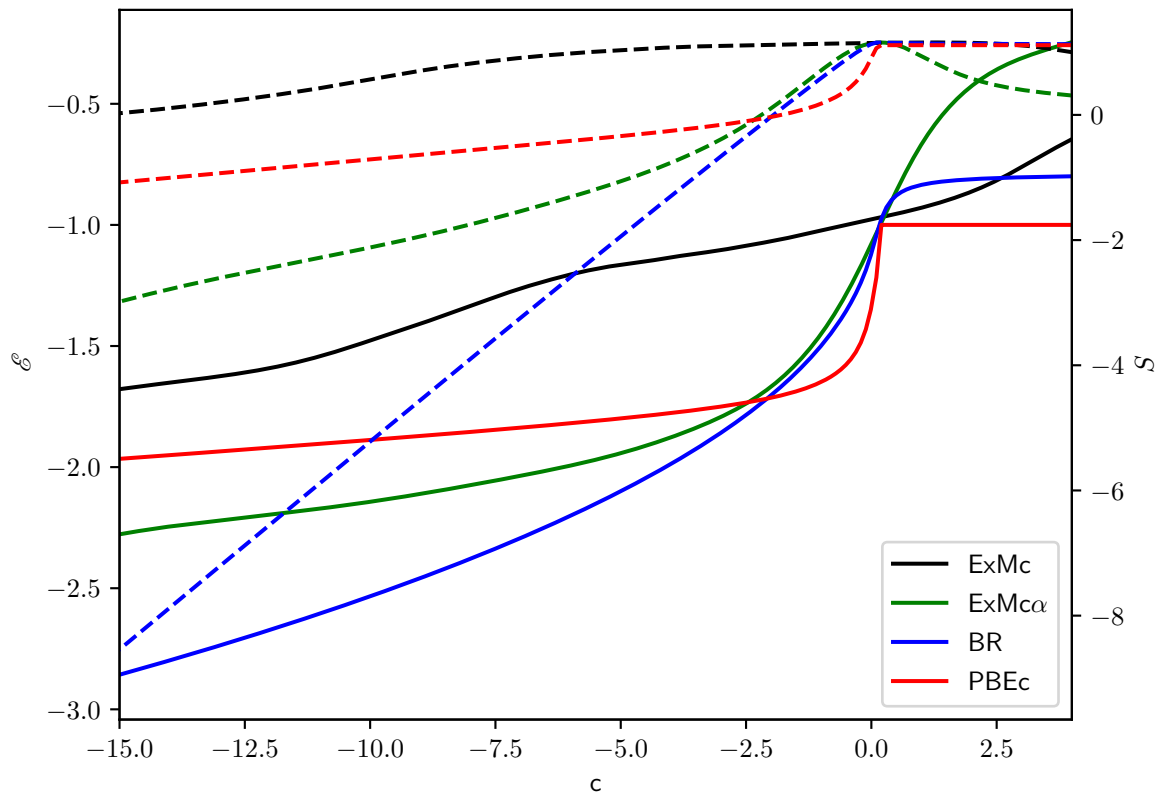


Fig. 3.3. Exchange energies per particle in reduced units, $\epsilon_X/|\epsilon_X^{\text{LDA}}| = \mathcal{E}_X$ (full lines) and entropy (dashed lines) of various functionals defined in the text versus the reduced curvature c .

In Table III-1 we provide E_X for a few atoms. ExMc is not negative enough for all atoms, similar to LDA. This is not too surprising in light of the weak dependence of ExMc on c . The benefit that can be achieved through the introduction of c is only realized if C-III is imposed over a range of u values around $u = 0$. The parameter α might be determined empirically or from constraints (such as $d^4\rho_X(\mathbf{r}, u = 0)/du^4$ [49]), here we do not pursue this further, rather we turn our attention to a procedure for modifying given X holes through the addition of further conditions, including the consideration of correlation effects.

Tab. III-1. Exchange energies for atoms in Hartree units. All approximations have been implemented self-consistently with respect to variations of the orbitals. ExMc and ExMc α are the functionals generated by the machine learning algorithm (ExMachina) proposed here. LDA denotes the local approximation derived from the homogeneous electron gas and BR represents the Becke-Roussel approximation. "Exact" stands for self-consistent Hartree-Fock exchange. MAE abbreviates mean absolute error.

Atom	ExMc	ExMc α	LDA	BR	Exact
He	-0.843	-0.993	-0.855	-1.031	-1.026
Ne	-10.690	-11.833	-10.959	-12.128	-12.114
Ar	-27.061	-29.426	-27.782	-29.990	-30.186
Kr	-85.991	-91.748	-88.464	-92.678	-93.858
MAE	3.2	0.8	2.3	0.3	-

3.3. Evolving a given X or XC hole through machine generated factors

Starting from a given X or XC hole which satisfies n conditions ($\rho_{X(C)}^n(\mathbf{r}, u)$), additional ones can be added through the factorization ansatz [25–29] $\rho_{X(C)}^{n+1}(\mathbf{r}, u) = f^n(\mathbf{r}, u)\rho_{X(C)}^n(\mathbf{r}, u)$. To consider an example, building on the curvature dependent $\rho_X^{\text{ExMc}}(u, \mathbf{c})$, or an established hole model such as $\rho_X^{\text{BR}}(u, \mathbf{c})$, we can incorporate further conditions and for the purpose of illustration, we impose a given energy density per particle ϵ ,

$$\rho_{X(C)}^{\text{ExMc}\epsilon}(u, \mathbf{c}, \epsilon) = f^{\text{ExMc}\epsilon}(u, \mathbf{c}, \epsilon)\rho_X(u, \mathbf{c}). \quad (3.11)$$

This scheme can for instance be used to reproduce a given exchange functional such as PBE or B88 [50] (C-V) and thus to obtain new X hole models from existing ones. Another application of interest is to impose the exchange plus static correlation energy density per particle $\epsilon_{\text{XS}}(\mathbf{r})$ as it is obtained from a Multi-Configuration Self-Consistent Field (MCSCF) calculation. Recently [29], we used $\rho_X^{\text{BR}}(u, \mathbf{c})$ to represent $\epsilon_{\text{XS}}(\mathbf{r})$ and to satisfy the additional constraint, we abandoned the normalization C-IV. Here we preserve C-IV, in fact all the conditions that were already imposed on $\rho_X^{\text{BR}}(u, \mathbf{c})$ previously. The factor $f^{\text{ExMc}\epsilon}(u, \mathbf{c}, \epsilon)$ is accordingly designed as a modulation of $\rho_X^{\text{BR}}(u, \mathbf{c})$ such that the model hole $f^{\text{ExMc}\epsilon}(u, \mathbf{c}, \epsilon)\rho_X^{\text{BR}}(u, \mathbf{c})$ satisfies the constraints C-I to C-V.

In addition to adjusting the constraints, we also replace the entropy in Eq. 2.3 to obtain,

$$\Lambda[f(\mathbf{c}, \epsilon)] = \int du (f(u, \mathbf{c}, \epsilon) - 1)^2 + \sum_{i=1}^5 \lambda_i (\mathbf{c}^{(i)}[f(\mathbf{c}, \epsilon)\rho_X^{\text{BR}}(\mathbf{c})] - \mathbf{c}^{(i)}), \quad (3.12)$$

where $\epsilon = \mathbf{c}^{(5)}$ and $\mathbf{c} = \mathbf{c}^{(3)}$. The first term in this expression is the least square deviation of $f(u, \mathbf{c}, \epsilon)$ with respect to one, thus we require $f(u, \mathbf{c}, \epsilon)$ to represent a minimal adjustment of the underlying X hole. $\Lambda[f(\mathbf{c}, \epsilon)]$ has the desirable property of being directly applicable to introduce correlation. This is because $f(u, \mathbf{c}, \epsilon)$ can be negative as well as positive, as required to model $\rho_{XC}(\mathbf{r}, u)$. There is no obvious way to apply the maximum entropy approach to generate approximations to $\rho_{XC}(\mathbf{r}, u)$. Making $\Lambda[f(\mathbf{c}, \epsilon)]$ stationary with respect to f and the λ_i defines the factor $f^{\text{ExMc}\epsilon}(u, \mathbf{c}, \epsilon)$. $\Lambda[f(\mathbf{c}, \epsilon)]$ depends on \mathbf{c} and ϵ through f and through the constraints. To obtain $f^{\text{ExMc}\epsilon}(u, \mathbf{c}, \epsilon)$, it is expressed in reduced units y, c , and \mathcal{E} , introduced previously, $\mathbf{f}^{\text{ExMc}\epsilon}(y, c, \mathcal{E}) = f^{\text{ExMc}\epsilon}(u, \mathbf{c}, \epsilon)$. Then $\mathbf{f}^{\text{ExMc}\epsilon}(y, c, \mathcal{E})$ is represented through a NN, as sketched in Fig. 3.1. This NN is very similar to the NN employed for the X hole. The differences being that there are now three neurons in the first layer, representing the reduced variables. Furthermore, the activation function of the first three layers is hyperbolic tangent (\tanh) and the last one is softplus. This choice of activation functions is arrived at through trial and error while observing the convergence of the PyTorch optimization procedure.

Again, we have to translate the functional in Eq. 3.12 into a loss function and we follow the established pattern of using the penalty method,

$$\begin{aligned} \Lambda^P[f] = & \sum_{j,k}^{N,M} \left(\int du (f(u, \mathbf{c}_j, \epsilon_k) - 1)^2 \right. \\ & + \sigma(1 - \delta_{\epsilon_k, \epsilon_0}) \left(\sum_{i=1,2,4} (\mathfrak{C}_i[f(u, \mathbf{c}_j, \epsilon_k) \rho_X(\mathbf{c}_j, \epsilon_k)] - \mathbf{c}_i)^2 \right. \\ & \left. \left. + (\mathfrak{C}_3[f(u, \mathbf{c}_j, \epsilon_k) \rho_X(\mathbf{c}_j, \epsilon_k)] - \mathbf{c}_j)^2 + (\mathfrak{C}_5[f(u, \mathbf{c}_j, \epsilon_k) \rho_X(\mathbf{c}_j, \epsilon_k)] - \epsilon_k)^2 \right) \right). \end{aligned} \quad (3.13)$$

In this expression, in addition to the curvature, the energy per particle $\epsilon = \mathcal{E}\epsilon^{\text{LDA}}$ is discretized on a grid for \mathcal{E} , with 10 points ranging from 0.8 \mathcal{E}_0 to 1.2 \mathcal{E}_0 , where $\mathcal{E}_0 = \epsilon_0/\epsilon_X^{\text{LDA}}$. \mathcal{E}_0 is given by the reduced X energy per particle of the underlying X hole $\rho_X(u, \mathbf{c})$. Here $\mathcal{E}_0 = \epsilon_X^{\text{BR}}/\epsilon_X^{\text{LDA}}$. The grid for the curvature c with 18 uniformly distributed points ranges from -4 to 5 ; this c range is sufficient for the purpose of illustrating the model for the stretched H_2 molecule below. The optimization procedure of the NN is the same as used for J_X^{ExMc} . Furthermore, to guaranty coherence of the models, the factor $(1 - \delta_{\epsilon_k, \epsilon_0})$ is introduced in Eq. 3.13. It ensures that the function $f(u, \mathbf{c}_j, \epsilon_0)$ is identical to one if $\epsilon = \epsilon_0$.

To illustrate the resulting $\rho_X^{\text{ExMce}}(u, c, \epsilon)$, in Fig. 3.5 we consider stretched H_2 and compare reduced hole models to the exact reduced X hole. The model and exact holes are calculated with Hartree-Fock orbitals using a 6-31g** basis set. Further technical details are provided in Sec. 4.2. Since the starting point is an atomic size BR X hole, the challenge is to model a molecular X hole on an atomic scale set by the range of J_X^{BR} . The reference points that we consider are along the bond axis of H_2 , they are indicated in Fig. 3.4. First, we focus on the bond midpoint (A) and compare the various X holes in Fig. 3.5. At A, the exact X hole is atomic like (due to the symmetry around the reference position) and the imposition of \mathcal{E} provides an improvement in J^{ExMce} over J_X^{BR} . Similarly, the relevant integrands for the energy (yJ) and the normalization (y^2J) are also improved. The exact exchange hole becomes de-localized as one moves from A to B where there are now two minima in the exact hole that cannot be represented by BR. J^{ExMce} does exhibit two minima, albeit on an atomic length scale. The diminished de-localization of J^{ExMce} , as compared to the exact hole, is visible in the lowest plot of the B column where the radial hole densities are compared. At position B, the factor $\mathbf{f}^{\text{ExMce}}(y, c, \mathcal{E})$ reduces the absolute value of the exchange energy per particle of J_X^{BR} . On the other hand, the normalization of the X hole is preserved. The former constraint is satisfied by reducing the hole depth at intermediate range and the latter condition requires an increased depth at longer range; these two effects combined produce the oscillation in J^{ExMce} . In general, for a poly-atomic molecule, there will be numerous minima in the exact X hole and there is little hope that a model can reproduce them at the correct distance. From a point of view of adding correlation effects through an additional

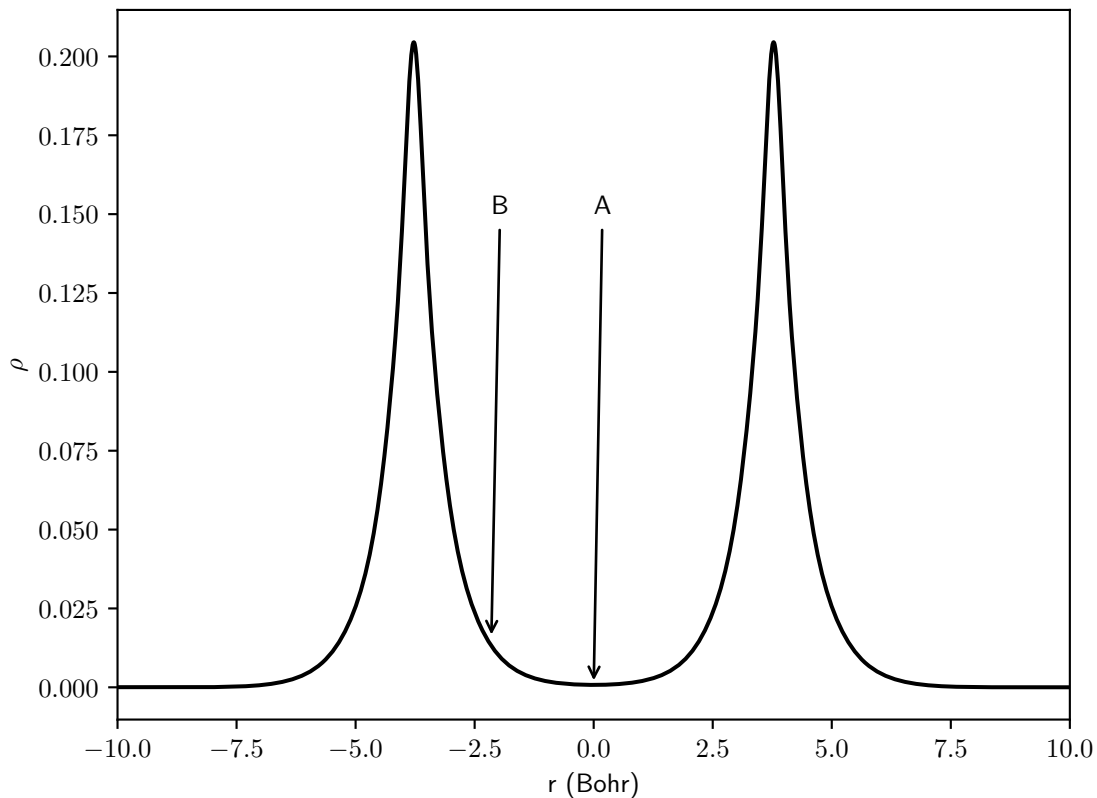


Fig. 3.4. The electron density (ρ) of stretched H_2 (bond length $R = 7.56$ Bohr) along the molecular axis. A and B label the positions at which the X holes are compared.

correlation factor, it is desirable [25, 26] to have a simple, predictable hole model such as provided by J^{ExMcc} rather than the exact hole or a model hole that reproduces the complex de-localization of the exact one. It is, however, crucial for the model X hole to reproduce the exact ϵ_X so that it can be recovered in high density and one-electron regions where it represents the correct value of the XC energy [28].

4. Technical Details

4.1. Neural networks

The reduced holes J_X^c and the factor \mathbf{f}^{ce} developed here are represented by feed forward NNs with three hidden layers containing each 100 neurons as sketched in Fig. 3.1, an architecture similar to the one employed by Nagai and coworkers [13]. In these NNs, the mapping between the input \mathbf{x} ($\mathbf{x} = (y, c)$ or $\mathbf{x} = (y, c, \mathcal{E})$) and the output ω (J_X^c or \mathbf{f}^{ce}) and is $\omega = h_4(h_3(h_2(h_1(\mathbf{x}))))$. The layers h_i defined in terms of their activation functions

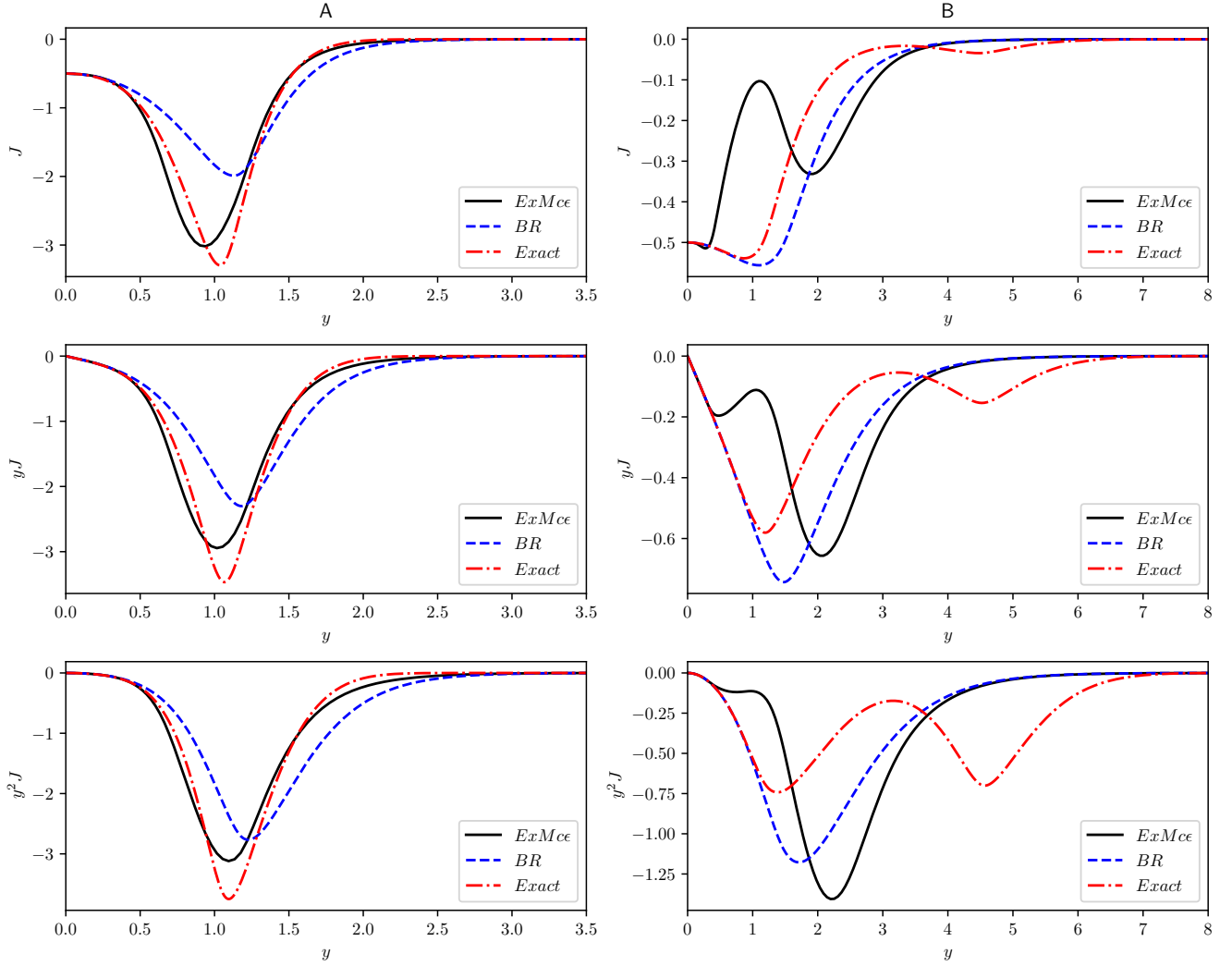


Fig. 3.5. Comparison of the reduced X holes of $\rho_X^{\text{ExMce}}(u,c,\epsilon)$ and $\rho_X^{\text{BR}}(u,c)$ with the exact reduced X hole. The system considered is the H_2 molecule (bond length $R = 7.56$ Bohr). In addition to the reduced holes, the integrands of the first and second moment are compared as well. The reference positions A (left column) and B (right column) are as described in Fig. 3.4.

ϕ_i and their sets of weights W_i and biases b_i are $h_i(\mathbf{v}) = \phi_i(\mathbf{W}_i\mathbf{v} + \mathbf{b}_i)$. For an input \mathbf{x} of dimension N , the weight matrices have the following dimensions: $\dim\mathbf{W}_1 = 100 \times N$, $\dim\mathbf{W}_2 = 100 \times 100$, $\dim\mathbf{W}_3 = 100 \times 100$, $\dim\mathbf{W}_4 = 1 \times 100$. The bias vectors have the dimensions: $\dim\mathbf{b}_1 = \dim\mathbf{b}_2 = \dim\mathbf{b}_3 = 100$ and $\dim\mathbf{b}_4 = 1$. The matrices \mathbf{W}_i and vectors \mathbf{b}_i are determined by the minimization of the loss functions defined in Eqs. 3.9 and 3.13.

For the two NNs described, the Adam optimizer [51] was used with a starting learning rate of 0.0001. The learning rate was lowered if the loss function was oscillating during optimization.

4.2. Evaluation of spherically-averaged exchange holes

Here we describe the procedure employed to obtain the spherically-averaged, exact exchange holes presented in figure 3.4. In the spin-unpolarized case, the exact exchange hole can be expressed in terms of the Hartree-Fock or KS first-order reduced density matrix $\gamma(\mathbf{r},\mathbf{r}') = 2 \sum_i^{occ} \psi_i(\mathbf{r})\psi_i(\mathbf{r}')$, where ψ_i is a Hartree-Fock or KS spatial orbital,

$$\rho_X(\mathbf{r},\mathbf{r}') = -\frac{\gamma(\mathbf{r},\mathbf{r}')\gamma(\mathbf{r}',\mathbf{r})}{2\rho(\mathbf{r})}. \quad (4.1)$$

Expressing $\gamma(\mathbf{r},\mathbf{r}')$ in a Gaussian orbital basis $\{\chi_\mu\}$, we find

$$\gamma(\mathbf{r},\mathbf{r}')\gamma(\mathbf{r}',\mathbf{r}) = \sum_{\mu\nu\kappa\lambda} P_{\mu\nu}P_{\kappa\lambda}\chi_\mu(\mathbf{r})\chi_\kappa(\mathbf{r})\chi_\nu(\mathbf{r}')\chi_\lambda(\mathbf{r}'), \quad (4.2)$$

where $P_{\mu,\nu}$ is the representation of the density matrix in the atomic basis. Defining $F_\nu(\mathbf{r}) = \sum_\mu P_{\mu\nu}\chi_\mu(\mathbf{r})$ and replacing \mathbf{r}' by $\mathbf{r} + \mathbf{u}$ yields

$$\rho_X(\mathbf{r},\mathbf{r} + \mathbf{u}) = -\frac{\sum_{\nu\lambda} F_\nu(\mathbf{r})F_\lambda(\mathbf{r})\chi_\nu(\mathbf{r} + \mathbf{u})\chi_\lambda(\mathbf{r} + \mathbf{u})}{2\rho(\mathbf{r})}. \quad (4.3)$$

Subsequently, the spherical average ($\int \frac{d\Omega_u}{4\pi}$) over orientations of \mathbf{u} is calculated,

$$\rho_X(\mathbf{r},u) = \int \frac{d\Omega_u}{4\pi} \rho_X(\mathbf{r},\mathbf{r} + \mathbf{u}) = \frac{-1}{2\rho(\mathbf{r})} \sum_{\nu\lambda} F_\nu(\mathbf{r})F_\lambda(\mathbf{r}) \int \frac{d\Omega_u}{4\pi} \chi_\nu(\mathbf{r} + \mathbf{u})\chi_\lambda(\mathbf{r} + \mathbf{u}). \quad (4.4)$$

The integrals $\int \frac{d\Omega_u}{4\pi} \chi_\nu(\mathbf{r} + \mathbf{u})\chi_\lambda(\mathbf{r} + \mathbf{u})$ are evaluated using the implementation described in earlier work [52].

5. Conclusions

The development of approximations to E_{XC} is a non-systematic process and there are at least as many approaches to this task as there are researchers in this field. In the present work, we take steps towards a formalized approximation strategy (ExMachina) that exploits ML algorithms which recently became available. NNs are highly flexible tools and changes in the NN architecture, such as modifying the activation function, can lead to distinct results. However, we show that physically meaningful models for XC holes can be obtained through machine learning. It is reasonable to expect that as the number of physical constraints applied increases, the degree of non-uniqueness in the hole models is reduced.

We illustrate the ExMachina generation of approximations by constructing the X holes ExMc and ExMc α . Both of these approximations provide further insight into curvature-dependent models to E_X and it becomes apparent that the range over which we impose a condition is a determining factor. In addition to exploiting physical constraints, as it is done here, it is of course also possible to include empirical information through corresponding terms added to the functional $L[\rho_X]$.

Employing the factorization ansatz, $\rho_{X(C)}^{n+1}(\mathbf{r}, u) = f^n(\mathbf{r}, u)\rho_{X(C)}^n(\mathbf{r}, u)$, and building on an existing hole model, we show how to use ExMachina to further refine and adapt a given approximation. As an illustration, we develop the NN representation of f^n that results in the first X hole model that reproduces a given ϵ_X and respects normalization. We can also adapt f^n to include correlation (beyond static correlation). To this end, f^n can be constrained to reproduce an approximate on-top $\rho_{XC}(\mathbf{r}, u = 0)$ value of a correlated hole. The on-top value, through the coulomb cusp condition, also determines $d\rho_{XC}(\mathbf{r}, u = 0)/du$ and this provides an additional constraint. In our previous work, we employed this strategy in the construction of, for instance, the CFXStatic [29], albeit through traditional means.

As shown with ExMachina, ML offers enormous flexibility to advance representations of XC holes and to develop self-consistent implementations of the resulting approximations to E_{XC} . In the present work we provide two distinct variants (Eq. 3.8 and Eq. 3.12) for the transformation of physical constraints into new approximations. Instead of improving approximations directly, with ExMachina we pursue theories and algorithms that generate improved approximations. The construction of approximations is as much an art as a science and ML can help to move it towards the latter.

Acknowledgments

The financial support through *The Natural Sciences and Engineering Research Council of Canada (NSERC)* is gratefully acknowledged. Computations were made on the super-computer Béluga, managed by Calcul Québec and Compute Canada. The operation of this supercomputer is funded by the Canada Foundation for Innovation (CFI), Ministère de l'Économie et de l'Innovation du Québec (MEI) and le Fonds de recherche du Québec (FRQ).

Author declarations

Conflict of Interest

The authors have no conflicts to disclose.

Data Availability Statement

The data that supports the findings of this study are available within the article.

References

- [1] W. Kohn and L. J. Sham, *Physical Review* **140**, A1133 (1965).
- [2] R. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules (New York, 1989)*. (Oxford University Press, New York, 1989).
- [3] R. Dreizler and E. Gross, *Density-Functional Theory (Springer, Berlin, 1990)* (Springer, 1990).
- [4] A. D. Becke, *The Journal of Chemical Physics* **140**, 18A301 (2014).
- [5] B. Kalita, L. Li, R. J. McCarty, and K. Burke, *Accounts of Chemical Research* **54**, 818 (2021).
- [6] S. Manzhos, *Machine Learning: Science and Technology* **1**, 013002 (2020).
- [7] J. A. Keith, V. Vassilev-Galindo, B. Cheng, S. Chmiela, M. Gastegger, K.-R. Müller, and A. Tkatchenko, *Chemical Reviews* **121**, 9816 (2021).
- [8] K. Ryczko, D. A. Strubbe, and I. Tambllyn, *Physical Review A* **100**, 022512 (2019).
- [9] M. Bogojeski, L. Vogt-Maranto, M. E. Tuckerman, K.-R. Müller, and K. Burke, *Nature Communications* **11**, 5223 (2020).
- [10] F. Brockherde, L. Vogt, L. Li, M. E. Tuckerman, K. Burke, and K.-R. Müller, *Nature Communications* **8**, 872 (2017).
- [11] J. C. Snyder, M. Rupp, K. Hansen, K.-R. Müller, and K. Burke, *Physical Review Letters* **108**, 253002 (2012).
- [12] R. Nagai, R. Akashi, S. Sasaki, and S. Tsuneyuki, *The Journal of Chemical Physics* **148**, 241737 (2018).
- [13] R. Nagai, R. Akashi, and O. Sugino, *npj Computational Materials* **6**, 43 (2020).
- [14] L. Li, S. Hoyer, R. Pederson, R. Sun, E. D. Cubuk, P. Riley, and K. Burke, *Physical Review Letters* **126**, 036401 (2021).
- [15] M. F. Kasim and S. M. Vinko, *Phys. Rev. Lett.* **127**, 126403 (2021).
- [16] S. Dick and M. Fernandez-Serra, *Nature Communications* **11**, 3509 (2020).
- [17] J. Schmidt, C. L. Benavides-Riveros, and M. A. L. Marques, *The Journal of Physical Chemistry Letters* **10**, 6425 (2019).
- [18] J. Gedeon, J. Schmidt, M. J. Hodgson, J. Wetherell, C. L. Benavides-Riveros, and M. A. L. Marques, *Machine Learning: Science and Technology* (2021).
- [19] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, in *Advances in Neural Information Processing Systems*, Vol. 32, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Curran Associates, Inc., 2019).
- [20] I. Lagaris, A. Likas, and D. Fotiadis, *Computer Physics Communications* **104**, 1 (1997).
- [21] H. Nakanishi and M. Sugawara, *Chemical Physics Letters* **327**, 429 (2000).

- [22] M. Sugawara, *Computer Physics Communications* **140**, 366 (2001).
- [23] V. Subramanian, *Deep Learning with PyTorch: A Practical Approach to Building Neural Network Models Using PyTorch* (Packt Publishing Ltd, 2018).
- [24] H. Bahmann and M. Ernzerhof, *The Journal of Chemical Physics* **128**, 234104 (2008).
- [25] J. Přecechtělová, H. Bahmann, M. Kaupp, and M. Ernzerhof, *The Journal of Chemical Physics* **141**, 111102 (2014).
- [26] J. Přecechtělová, H. Bahmann, M. Kaupp, and M. Ernzerhof, *The Journal of Chemical Physics* **143**, 144102 (2015).
- [27] R. Wang, Y. Zhou, and M. Ernzerhof, *The Journal of Chemical Physics* **150**, 084107 (2019).
- [28] R. Wang, Y. Zhou, and M. Ernzerhof, *The Journal of Chemical Physics* **151**, 194102 (2019).
- [29] P.-O. Roy, E. Cuierrier, and M. Ernzerhof, *The Journal of Chemical Physics* **152**, 211101 (2020).
- [30] A. Savin, in *Recent Developments and Applications of Modern Density Functional Theory*, Theoretical and Computational Chemistry, Vol. 4, edited by J. Seminario (Elsevier, 1996) pp. 327 – 357.
- [31] T. Leininger, H. Stoll, H.-J. Werner, and A. Savin, *Chemical Physics Letters* **275**, 151 (1997).
- [32] H. Iikura, T. Tsuneda, T. Yanai, and K. Hirao, *The Journal of Chemical Physics* **115**, 3540 (2001).
- [33] J. Toulouse, F. Colonna, and A. Savin, *Physical Review A* **70**, 062505 (2004).
- [34] A. V. Krukau, G. E. Scuseria, J. P. Perdew, and A. Savin, *The Journal of Chemical Physics* **129**, 124103 (2008).
- [35] T. M. Henderson, B. G. Janesko, and G. E. Scuseria, *The Journal of Chemical Physics* **128**, 194105 (2008).
- [36] M. Modrzejewski, M. Hapka, G. Chalasinski, and M. M. Szczesniak, *Journal of Chemical Theory and Computation* **12**, 3662 (2016).
- [37] J. Tao and Y. Mo, *Physical Review Letters* **117**, 073001 (2016).
- [38] J. Tao, I. W. Bulik, and G. E. Scuseria, *Physical Review B* **95**, 125115 (2017).
- [39] M. Ernzerhof and J. P. Perdew, *The Journal of Chemical Physics* **109**, 3313 (1998).
- [40] A. D. Becke and M. R. Roussel, *Physical Review A* **39**, 3761 (1989).
- [41] M. Kelbert, I. Stuhl, and Y. Suhov, *Modern Stochastics: Theory and Applications* **4**, 233 (2017).
- [42] R. Stewart and S. Ermon, in *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 31 (2017).
- [43] J. Antolín, A. Zarzo, and J. C. Angulo, *Physical Review A* **50**, 240 (1994).

- [44] A. E. Smith and D. W. Coit, Penalty functions handbook of evolutionary computation (Oxford University Press and Institute of Physics Publishing, 1997) Chap. Section C 5.2.
- [45] E. Proynov, Z. Gan, and J. Kong, Chemical Physics Letters **455**, 103 (2008).
- [46] J. P. Perdew, K. Burke, and M. Ernzerhof, Physical Review Letters **77**, 3865 (1996).
- [47] Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, S. Wouters, and G. K.-L. Chan, WIREs Computational Molecular Science **8**, e1340 (2018).
- [48] Q. Sun, X. Zhang, S. Banerjee, P. Bao, M. Barbry, N. S. Blunt, N. A. Bogdanov, G. H. Booth, J. Chen, Z.-H. Cui, J. J. Eriksen, Y. Gao, S. Guo, J. Hermann, M. R. Hermes, K. Koh, P. Koval, S. Lehtola, Z. Li, J. Liu, N. Mardirossian, J. D. McClain, M. Motta, B. Mussard, H. Q. Pham, A. Pulkin, W. Purwanto, P. J. Robinson, E. Ronca, E. Sayfutyarova, M. Scheurer, H. F. Schurkus, J. E. T. Smith, C. Sun, S.-N. Sun, S. Upadhyay, L. K. Wagner, X. Wang, A. White, J. D. Whitfield, M. J. Williamson, S. Wouters, J. Yang, J. M. Yu, T. Zhu, T. C. Berkelbach, S. Sharma, A. Sokolov, and G. K.-L. Chan, The Journal of Chemical Physics **153**, 024109 (2020).
- [49] R. Wang, Y. Zhou, and M. Ernzerhof, Physical Review A **96**, 022502 (2017).
- [50] A. D. Becke, Physical Review A **38**, 3098 (1988).
- [51] D. P. Kingma and J. Ba, arXiv:1412.6980 [cs] (2017).
- [52] H. Antaya, Y. Zhou, and M. Ernzerhof, Physical Review A **90**, 1 (2014).

Troisième article.

The fourth-order expansion of the exchange hole and neural networks to construct exchange-correlation functionals

par

Etienne Cuierrier¹, Pierre-Olivier Roy¹, Rodrigo Wang² and Matthias Ernzerhof¹

(¹) Département de Chimie, Université de Montréal, C.P. 6128 Succursale A, Montréal, Québec H3C 3J7, Canada

(²) 1QB Information Technologies (1QBit), Vancouver, British Columbia 4B1, Canada

Cet article a été publié dans *Journal of Chemical Physics*, 157, 171103 (2022);
<https://doi.org/10.1063/5.0122761>.

Les principales contributions de Étienne Cuierrier à cet article sont présentées. J'ai fait la construction des modèles, le code pour entraîner les réseaux de neurones et les calculs pour les systèmes chimiques. J'ai aussi contribué à l'écriture de l'article et à l'analyse des résultats. Pierre-Olivier Roy a participé au code pour calculer le quatrième terme de l'expansion en série de Taylor du trou d'échange exact. Rodrigo Wang a participé aux discussions.

Cet article concerne le développement de nouvelles approximations de l'énergie d'échange-corrélation en théorie de la fonctionnelle de la densité. Dans une étude de 2017 par notre groupe de recherche, nous avons montré que le quatrième terme de l'expansion en série de Taylor du trou d'échange ($T(\mathbf{r})$) est une propriété qui permet d'améliorer la description du trou d'échange des molécules. Cependant, construire des approximations à l'énergie électronique avec $T(\mathbf{r})$ est une tâche difficile. Avec l'expertise que nous avons développée en apprentissage automatique dans le deuxième projet de cette thèse, nous nous sommes donc intéressés à construire des approximations à l'énergie d'échange-corrélation avec des réseaux de neurones artificiels et $T(\mathbf{r})$.

Notre étude montre que les modèles que nous avons construits, en utilisant des NN et $T(\mathbf{r})$, permettent de créer une fonctionnelle dont la qualité des prédictions s'approche de celle d'une fonctionnelle hybride pour le calcul d'énergies d'atomisation, de barrières d'activation et de prédiction de la densité électronique. Notre approche à l'avantage que $T(\mathbf{r})$ a une complexité computationnelle plus faible que le calcul de l'échange d'une fonctionnelle hybride. Ainsi, notre modèle s'approche de la qualité d'une fonctionnelle hybride, en ayant potentiellement un temps de calcul plus faible.

ABSTRACT. The curvature Q_σ of the spherically-averaged exchange (X) hole $\rho_{X,\sigma}(\mathbf{r},u)$ is one of the crucial variables for the construction of approximations to the exchange-correlation (XC) energy of Kohn-Sham theory, the most prominent example being the Becke-Roussel model [A.D. Becke, M.R. Roussel, Phys. Rev. A 39, 3761 (1989)]. Here, we consider the next higher non-zero derivative of the spherically-averaged X hole, the fourth-order term T_σ . This variable contains information about the non-locality of the X hole and we employ it to approximate hybrid functionals, eliminating the sometimes demanding calculation of the exact X energy. The new functional is constructed using machine learning; having identified a physical correlation between T_σ and the non-locality of the X-hole, we employ a neural network to express this relation. While we only modify the X functional of the Perdew-Burke-Ernzerhof functional [J.P. Perdew, K. Burke, M. Ernzerhof, Phys. Rev. Lett. 77, 3865 (1996)], a significant improvement over this method is achieved.

Keywords: Density functional theory, Kohn-Sham, Exchange-correlation energy, machine learning, neural networks

1. Introduction

The Kohn-Sham (KS) [1–6] approach considers a system of fictitious non-interacting electrons with the same electronic density as the real system. In KS theory, the exchange-correlation (XC) energy (E_{XC}), accounting for the electron-electron interaction, must be approximated in a computationally affordable manner. The approximations to E_{XC} are often classified with Jacob's ladder [7], where they are stacked according to increasing computational cost and complexity. On the lowest rung, [2] the XC energy is represented using the

local spin-density ($\rho_\sigma(\mathbf{r})$) approximation (LSDA). Although this method is often successful in solid state physics, it is of limited use for chemistry. The generalized-gradient approximation (GGA) is the next rung on Jacob’s ladder, accurate enough for many problems in chemistry. In addition to $\rho_\sigma(\mathbf{r})$, its gradient $\nabla\rho_\sigma(\mathbf{r})$ is also considered. A popular GGA is the Perdew-Burke-Ernzerhof functional (PBE) [8], which is the basis for the development presented in this work. On the next higher rung, the kinetic energy density $\tau_\sigma(\mathbf{r})$ is added, the culmination of this type of approximation is perhaps the SCAN meta-GGA (MGGA) functional [9]. On the next rung of the ladder, approximations also use the occupied orbitals to include, for instance, a fraction of the exact X energy (e.g. B3LYP [10, 11] or PBEh [8, 12, 13]). For the purpose of this article, we stop at the hybrid rung of Jacob’s ladder.

Hybrid functionals are extremely successful in chemistry [5, 6], their fraction of exact X energy compensates for the over-localization of the XC hole of semi-local DFT functionals [14]. While hybrids are very popular, there has been work to replace them [6, 15] since the calculation of exact X is often the computational bottleneck, in particular for extended systems. Here we address this issue and we propose to use a new variable in the construction of functionals, reducing the computational cost but still providing an improvement upon PBE. In detail, we consider the spherically-averaged X hole $\rho_{X,\sigma}(\mathbf{r},u)$ [16] as a function of the reference point \mathbf{r} and the electron-electron separation u . Then we employ the series expansion of the X hole [17–19]

$$\rho_{X,\sigma}(\mathbf{r},u) = -\rho_\sigma(\mathbf{r}) - Q_\sigma(\mathbf{r})u^2 - T_\sigma(\mathbf{r})u^4 + \dots \quad (1.1)$$

to model the X contribution to PBE hybrid (PBEh), i.e.,

$$E_X^{PBEh} = \frac{1}{4}E_X^{exact} + \frac{3}{4}E_X^{PBE}. \quad (1.2)$$

Since the X hole is invariant w.r.t. orbital rotations, the same holds for its series expansion. In previous work [19], we already explored T_σ for the construction of X-hole models. The Becke-Roussel (BR) X hole [18] correctly reproduces the second-order expansion of the X hole and we tried to include T_σ in X hole models, we were unable, however, to find a flexible enough function to reproduce the second and fourth-order terms. In addition, we found a strong basis set dependence of T_σ in the core region, which also hampered the design of functionals. Therefore, until now, T_σ is a largely unexplored ingredient for the modeling of X and XC functionals. To solve our previous impasse, we resort to machine learning (ML), which became popular [15, 20–44] for the construction of approximations to E_{XC} . Here, since we want to eliminate the exact X energy, we use ML to reproduce the X energy of PBEh as closely as possible. The NN depends on the spin density, its gradient, Q_σ , and T_σ ; it is trained on only four molecules and the five atoms constituting those molecules. Since we have clearly identified, specific non-locality effects that we include into the newly designed functionals, no extensive datasets are needed for the ML procedure. PBE correlation is

added to the NN X functional and the resulting approximation yields an improved accuracy for atomization energies (AE), barrier heights (BH) and electronic densities, compared to PBE.

Related to this work is CIDER [15], which uses ML and non-local features to replace the calculation of exact exchange. Lei et al. [45] use ML to reproduce the B3LYP functional from the density. Janesko [46] developed the rung 3.5 functionals to replace exact exchange. Our methodology draws on the work of Nagai and coworkers [33], pioneering the usage of NN to improve XC functionals.

We begin the derivation of our approach with a summary of the X hole Taylor expansion.

2. The fourth-order expansion of the exchange hole

The XC energy is expressed in terms of the XC hole $\rho_{\text{XC}}(\mathbf{r}, \mathbf{r}')$ by

$$E_{\text{XC}} = \frac{1}{2} \int d\mathbf{r} d\mathbf{r}' \frac{\rho(\mathbf{r}) \rho_{\text{XC}}(\mathbf{r}, \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}. \quad (2.1)$$

The XC hole represents, for an electron at \mathbf{r} , the reduction of probability density to find another electron around it due to XC interactions. The XC hole can be separated in X and C holes,

$$\rho_{\text{XC}}(\mathbf{r}, \mathbf{r}') = \rho_{\text{X}}(\mathbf{r}, \mathbf{r}') + \rho_{\text{C}}(\mathbf{r}, \mathbf{r}') \quad (2.2)$$

or factorized using a C factor ansatz [47–51] pursued by our group

$$\rho_{\text{XC}}(\mathbf{r}, \mathbf{r}') = f_{\text{C}}(\mathbf{r}, \mathbf{r}') \rho_{\text{X}}(\mathbf{r}, \mathbf{r}'), \quad (2.3)$$

where $f_{\text{C}}(\mathbf{r}, \mathbf{r}')$ is the correlation factor. In this work, we will focus on the X hole. In the following, we consider only a single spin σ orientation, the total X hole is obtained as

$$\rho_{\text{X}}[\rho_{\sigma}, \rho_{\sigma'}](\mathbf{r}, u) = \sum_{\sigma} \frac{\rho_{\sigma}}{\rho} \rho_{\text{X}, \sigma}(\mathbf{r}, u). \quad (2.4)$$

The Coulomb potential has no dependence on the angle of $\mathbf{u} = \mathbf{r}' - \mathbf{r}$, we can thus perform a spherical average over all the orientations of \mathbf{u} ,

$$\rho_{\text{X}, \sigma}(\mathbf{r}, u) = \frac{1}{4\pi} \int_0^{2\pi} d\phi_u \int_0^{\pi} d\theta_u \sin(\theta_u) \rho_{\text{X}, \sigma}(\mathbf{r}, \mathbf{r} + \mathbf{u}). \quad (2.5)$$

The hole $\rho_{\text{X}, \sigma}(\mathbf{r}, u)$ has various known physical properties [16], e.g. it is negative,

$$\rho_{\text{X}, \sigma}(\mathbf{r}, u) \leq 0, \quad (2.6)$$

normalized

$$\int_0^{\infty} du 4\pi u^2 \rho_{\text{X}, \sigma}(\mathbf{r}, u) = -1, \quad (2.7)$$

and its on-top value is

$$\rho_{\text{X}, \sigma}(\mathbf{r}, u = 0) = -\rho_{\sigma}(\mathbf{r}). \quad (2.8)$$

The series expansion of the X hole up to second order was analyzed and exploited by Becke [17, 18], and extended to fourth order by our group [19],

$$\rho_{X,\sigma}(\mathbf{r},u) = -\rho_\sigma(\mathbf{r}) - Q_\sigma(\mathbf{r})u^2 - T_\sigma(\mathbf{r})u^4 + \dots, \quad (2.9)$$

where

$$Q_\sigma(\mathbf{r}) = \frac{1}{6} \left[\nabla^2 \rho_\sigma - 2\tau_\sigma + \frac{1}{2} \frac{(\nabla \rho_\sigma)^2}{\rho_\sigma} \right] \quad (2.10)$$

and

$$\begin{aligned} T_\sigma(\mathbf{r}) = & \frac{1}{5! \rho_\sigma} \left[2\rho_\sigma \sum_j^{occ} \psi_{j,\sigma} \nabla^4 \psi_{j,\sigma} + 4\nabla \rho_\sigma \cdot \sum_j^{occ} \psi_{j,\sigma} \nabla^3 \psi_{j,\sigma} \right. \\ & \left. + \frac{1}{2} (\nabla^2 \rho_\sigma - 2\tau_\sigma)^2 + 4 \sum_i^{occ} \psi_{i,\sigma} \psi_{i,\sigma}^{v6} \cdot \sum_j^{occ} \psi_{j,\sigma} \psi_{j,\sigma}^{v6} \right], \end{aligned} \quad (2.11)$$

with

$$\psi^{v6} = (\psi_{xx}, \psi_{yy}, \psi_{zz}, \sqrt{2}\psi_{xy}, \sqrt{2}\psi_{xz}, \sqrt{2}\psi_{yz}), \quad (2.12)$$

where the subscripts indicate a derivative with respect to the coordinates. In our previous work [19] we showed that the delocalized X-hole in a molecule cannot be captured by the series expansion of the X-hole to second order, but it requires at least a fourth-order expansion. Therefore, T_σ is a valuable new variable for the construction of X and XC functionals as demonstrated below. While the computation of T_σ is more tedious than that of Q_σ , both quantities only involve basis functions that have non-negligible values at a given point \mathbf{r} in the system, thus, they have the same computational scaling with system size. This has to be contrasted to the computation of exchange, a highly non-local quantity. Furthermore, the fourth derivative is straight-forward to calculate in a plane-wave basis often used for extended systems.

3. Machine learning to construct functionals drawing on the fourth-order expansion of the exchange hole

It is challenging to design an approximate X hole that reproduces Q_σ and T_σ [19], instead, we directly construct an X energy per particle ($\epsilon_{X,\sigma}(\mathbf{r})$) that depends on Q_σ and T_σ . Inspired by Refs. [33, 38], we use ML to add an additional degree of non-locality to semi-local functionals to approximate the X contribution of PBEh given in Eq. 1.2. The new X functional is defined through an enhancement factor $f^{NNx4}(\mathbb{S}_\sigma(\mathbf{r}), \mathbb{Q}_\sigma(\mathbf{r}), \mathbb{T}_\sigma(\mathbf{r}))$ for the PBE exchange energy density per particle ($\epsilon_{X,\sigma}^{PBE}(\mathbf{r})$). It is used in conjunction with the PBE approximation to the correlation energy (E_C^{PBE}),

$$E_{XC}^{NNx4} = E_C^{PBE} + \sum_\sigma \int d^3r \rho(\mathbf{r}) \epsilon_{X,\sigma}^{PBE}(\mathbf{r}) f^{NNx4}(\mathbb{S}_\sigma(\mathbf{r}), \mathbb{Q}_\sigma(\mathbf{r}), \mathbb{T}_\sigma(\mathbf{r})), \quad (3.1)$$

where we use the following variable transformations, omitting the \mathbf{r} -dependence for simplicity of notation,

$$\mathbb{S}_\sigma = \log(1 + s_\sigma(\mathbf{r})), \text{ with } s_\sigma(\mathbf{r}) = \frac{|\nabla\rho_\sigma(\mathbf{r})|}{2k_{f\sigma}(\mathbf{r})\rho_\sigma(\mathbf{r})}, \text{ and } k_{f\sigma}(\mathbf{r}) = (3\pi^2\rho_\sigma(\mathbf{r}))^{1/3}, \quad (3.2)$$

$$\mathcal{Q}_\sigma = Q_\sigma(\mathbf{r})/k_{f\sigma}^2(\mathbf{r})\rho_\sigma(\mathbf{r}), \quad (3.3)$$

$$\mathcal{T}_\sigma = T_\sigma(\mathbf{r})/k_{f\sigma}^4(\mathbf{r})\rho_\sigma(\mathbf{r}), \quad (3.4)$$

$$\mathbb{Q}_\sigma = \begin{cases} \log(1 + \mathcal{Q}_\sigma), & \text{if } \mathcal{Q}_\sigma \geq 0 \\ -\log(1 - \mathcal{Q}_\sigma), & \text{if } \mathcal{Q}_\sigma < 0, \end{cases} \quad (3.5)$$

$$\mathbb{T}_\sigma = \begin{cases} \log(1 + \mathcal{T}_\sigma), & \text{if } \mathcal{T}_\sigma \geq 0 \\ -\log(1 - \mathcal{T}_\sigma), & \text{if } \mathcal{T}_\sigma < 0. \end{cases} \quad (3.6)$$

These transformations ensure that the resulting X functional obeys the proper density scaling relation [52]. Furthermore, the range of values for the input variables can vary over orders of magnitude, thus it is advantageous for the NN optimization to use log functions of the variables. Large values of $|T_\sigma|$ are characteristic for the core region (see Fig. 4.S1 of the supplementary material) and for Gaussian basis sets, a large basis set dependence is observed in this region [19], which is, however, suppressed through the log transformation. In the supplementary material, we indeed show that the basis set dependence of the T_σ -dependent functionals is in line with other approximations such as PBE. Furthermore, the transformations employed also maintain the information about the sign of Q_σ and T_σ . The enhancement factor $f^{NNx4}(\mathbb{S}_\sigma, \mathbb{Q}_\sigma, \mathbb{T}_\sigma)$ is then represented by a NN, which on the scale of current NNs is quite small, it is described in Fig. 4.1. This simple NN is appropriate since we don't want to capture numerous features of extensive datasets, rather we force the NN to capture non-locality described by its variables.

Having introduced the NN, the loss function for the training of the NN is provided by

$$L = \lambda L_{AE_x} + L_{E_x} \quad (3.7)$$

with

$$L_{AE_x} = \frac{1}{4} \sum_i^4 (\Delta_{AE} E_{i;x}^{model} - \Delta_{AE} E_{i;x}^{PBEh})^2, \quad (3.8)$$

which is the error in the X contribution for the AE compared to PBEh (see Eq. 1.2). We use a value of $\lambda = 1.4$, to give more weight to the reaction energies, although using $\lambda = 1$ and $\lambda = 10$ does not significantly alter the results for the calculation of the validation set. In the training set, we consider four molecules: N_2 , H_2O , C_2H_4 , P_2 . The training set mostly consists of molecules with multiple bonds since PBE predicts the AE of molecules with single bonds quite well, but not those with multiple bonds. The training set is unusually small in

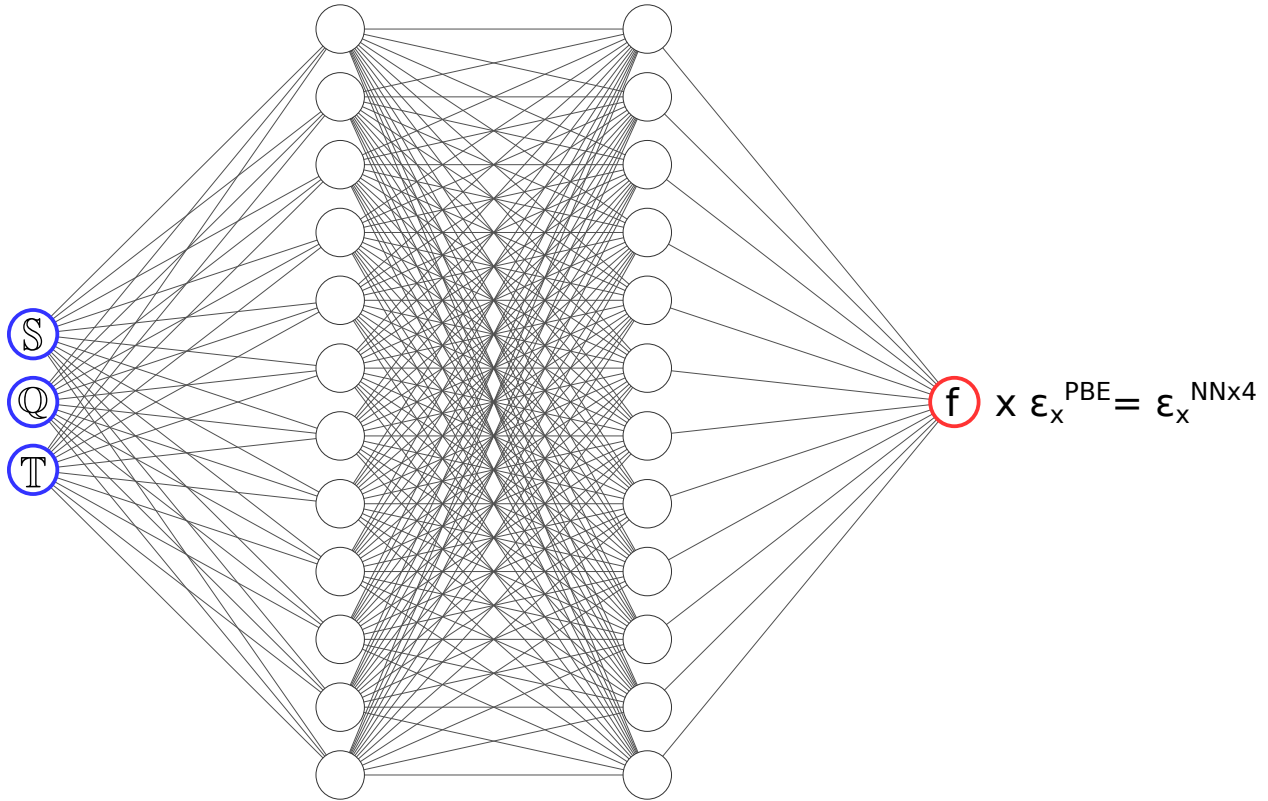


Fig. 4.1. The neural network used to represent the enhancement factor $f^{\text{NNx4}}(\mathbb{S}_\sigma, \mathbb{Q}_\sigma, \mathbb{T}_\sigma)$. This enhancement factor is then multiplied with the X energy per particle of the PBE approximation (ϵ_x^{PBE}) to yield a new X functional (ϵ_x^{NNx4}). The neural network consists of an input layer with the three variables \mathbb{S}_σ , \mathbb{Q}_σ , and \mathbb{T}_σ ; then there are two hidden layers that are twelve neurons wide; finally, there is a single output neuron. The figure is drawn using the tool of reference [53].

the context of ML approaches, this is done purposefully, since we merely want to capture an additional degree of non-locality which is represented in the dataset, rather than various kinds of unidentified effects. L_{Ex} corresponds to the error on the X energy for the four molecules and the five atoms (H, C, N, O, P) constituting the molecules,

$$L_{Ex} = \frac{1}{9} \sum_i^9 (E_{i;x}^{\text{model}} - E_{i;x}^{\text{PBEh}})^2. \quad (3.9)$$

Thus, we use a NN to model PBEh exchange, starting from PBE exchange. This approach has the advantage of simplicity, however, it does not satisfy constraints such as the homogeneous gas (HEG) limit. In the next section we will describe an alternative model which does satisfy this limit.

3.1. Alternative enhancement factor satisfying the homogeneous gas limit

Our model satisfying the HEG limit (NNHx4), is inspired by the equation used by Dick and Fernandez-Serra [38],

$$E_{XC}^{NNHx4} = E_C^{PBE} + \sum_{\sigma} \int d^3r \rho_{\sigma}(\mathbf{r}) f^{NNHx4}(s_{\sigma}, \mathcal{Q}_{\sigma}, \mathcal{T}_{\sigma}) \epsilon_{X,\sigma}^{PBE}(\mathbf{r}), \quad (3.10)$$

where

$$f^{NNHx4}(s_{\sigma}, \mathcal{Q}_{\sigma}, \mathcal{T}_{\sigma}) = \frac{2}{1 + e^{-g(s_{\sigma}, \mathcal{Q}_{\sigma}, \mathcal{T}_{\sigma})}} \quad (3.11)$$

and

$$\begin{aligned} g(s_{\sigma}, \mathcal{Q}_{\sigma}, \mathcal{T}_{\sigma}) &= \left(\tanh^2(1 + s_{\sigma}) + \tanh^2(\mathcal{Q}_{\sigma} - \mathcal{Q}_{\sigma}^{HEG}) + \tanh^2(\mathcal{Q}_{\sigma} - \mathcal{Q}_{\sigma}^{HEG}) \right) \\ &\times F^{NNHx4}(\mathcal{S}_{\sigma}, \mathcal{Q}_{\sigma}, \mathcal{T}_{\sigma}). \end{aligned} \quad (3.12)$$

In Eq. 3.11, the enhancement factor is constrained to vary in a range between zero and two. $F^{NNHx4}(\mathcal{S}_{\sigma}, \mathcal{Q}_{\sigma}, \mathcal{T}_{\sigma})$ in Eq. 3.12 is a NN with the same architecture as before and is optimized to the same loss function as NNx4. The hyperbolic tangent (tanh) functions in Eq. 3.12 ensure that in the limit of a HEG, the local spin-density (LSD) approximation is recovered. To this end, $\mathcal{Q}_{\sigma}^{HEG} = 1/5$ and $\mathcal{T}_{\sigma}^{HEG} = -3/175$ which are the variable values calculated from the analytic formula for the X hole of the HEG.

4. Assessing and analyzing NNx4 and related functionals

To study the impact of T_{σ} in the designed X functionals, we consider various molecular properties and compare our NN models to SCAN [9], PBEh, and MN15L [54]. To better understand the role of T_{σ} , we also optimize NNs using only s_{σ} and $(s_{\sigma}, \mathcal{Q}_{\sigma})$, respectively. The name of the model indicates the order of the inputs considered, i.e. NNx4 characterizes $f^{NNx4}(\mathcal{S}_{\sigma}, \mathcal{Q}_{\sigma}, \mathcal{T}_{\sigma})$, NNx2 indicates $f^{NNx2}(\mathcal{S}_{\sigma}, \mathcal{Q}_{\sigma})$, and NNx0 stands for $f^{NNx0}(\mathcal{S}_{\sigma})$.

The NNs are realized with the Pytorch [55] and Pytorch Lightning software, they are illustrated in Fig. 4.1 and the technical details as well as the details of the electronic structure calculations are provided in the supplementary material.

4.1. Exchange contribution to atomization energies

The ML approach aims at reproducing the X energy of PBEh and therefore, first we examine to what extent this objective is reached. In particular, the non-local contributions to X are of interest and AE are sensitive to those. Thus, we calculate the X contribution to the AE of the W4-11 subset [56] of the GMTKN55 database [57]. As shown in Tab. IV-1,

all of the models using NN considerably improve upon PBE. Despite the fact that NNx0

	PBE	NNx0	NNx2	NNx4	NNHx0	NNHx2	NNHx4
MAE	16.1	6.1	6.2	5.3	6.5	7.4	5.5

Tab. IV-1. Errors in the X contributions to atomization energies (ΔE_X) relative to PBEh, i.e. $\Delta E_X^{approx.} - \Delta E_X^{PBEh}$. The mean absolute errors are calculated for the 140 molecules of the W4-11 data set [56, 57] and are listed in kcal/mol. The functionals are evaluated with self-consistent PBE orbitals.

uses the same input (s_σ) as the PBE X-functional, it is fitted to reproduce the AE of PBEh, which explains the improvement for the MAE on the W4-11 subset compared to the reference functional PBE. As we argue below in Sec. 4.3, NNx0 is biased too much towards reproducing the AE, at the expense of other properties. NNHx0 has a slightly higher MAE than NNx0 for the X contribution to AE and does not seem to overfit significantly as we will also explain in Sec. 4.3. The functionals NNx2 and NNHx2, depending on Q_σ in addition to s_σ , yield a MAE slightly larger than NNx0 and NNHx0. From the construction of the PBE X-functional, see e.g. [16], we know that s_σ is employed to model the curvature of the X hole, i.e. Q_σ , implying that s_σ and Q_σ contain similar information; this explains why the addition of Q_σ does not yield an improvement compared to NN(H)x0.

The optimization of the NN parameters for NNHx2 shows an oscillating behavior resulting in a higher value of the loss function compared to the other NN functionals. For consistency with the other NN functionals, the optimization procedure is not modified to explore this problem. In any case, NNx4 and NNHx4 are the most important functionals for this study, since they include T_σ . Finally, we observe that the best results are obtained for NNx4 and NNHx4, this shows that T_σ includes useful information to model the X energy. Including the HEG limit in NNHx0, NNHx2, and NNHx4 leads to slightly higher MAEs compared to NNx0, NNx2, and NNx4.

4.2. Atomization energies and barrier heights

For the AE of the validation dataset shown in Tab. IV-2, we observe that all of our functionals improve the calculation of AE compared to PBE. NNx4 and NNHx4 reproduce best the results of PBEh for AE. Thus, the NN approach extracts useful information from the variable T_σ for the prediction of AE. For the dataset Tab. IV-2, adding the HEG limit leads to a slight increase of the MAEs, and higher mean errors (MEs) than observed for NNs that do not satisfy the HEG limit. The HEG has a localized X hole and tends to overestimate AE [58].

We also study AE for a larger data set (W4-11 [56]) as well as BH for the BH76 dataset [59] (Tab. IV-3). Again, all the NN functionals predict lower MAEs than PBE for both W4-11 and BH76 datasets. For our NN models, the MAE on the W4-11 generally diminishes as we

add variables, getting closer to PBEh’s results. NNx4 and NNHx4 have the best performance out of all the NN functionals, this again suggests that T_σ is a useful ingredient to improve the calculation of the X energy. None of our models have a lower MAE than the SCAN or MN15L methods, it is important to note that the NN functionals are obtained by only improving the X energy, the C energy is calculated from the GGA PBE. By construction, the NN functional performance is tied to PBEh.

Other than MN15L, NNx0 has the lowest MAE for the BHs, even lower than PBEh’s MAE. As discussed for the error in the prediction of electronic densities in Sec 4.3, this low MAE likely indicates overfitting, one would expect that, by construction, the NN functionals cannot significantly improve over the results of PBEh, since we are using it as a target functional of the NN optimization. NNx2 and NNx4, both yield MAEs very close to the PBEh functional for BH, thus adding T_σ as an input does not lead to a lower MAE for BH76. NNHx4 has a higher MAE than NNx4 for the BH76 dataset, but it has the lowest MAE among the NN models that satisfy the HEG limit.

4.3. Self-consistent electronic densities

One of the advantages of the NN representation is that we have access to the derivatives of the NN with respect to its input variables. This facilitates a self-consistent implementation, the details of which are described in the supplementary material. Unlike many previous models using NNs [33, 38], our loss function does not include a term to minimize the error in electronic densities. Instead, this metric is used as an indicator of an overall improvement compared to the reference functional (PBE). Worsening electron densities are often a sign for over-parametrization [63]. To compute the error on the densities, we use the following expression

$$\epsilon_{|n|} = \sum_i^N \frac{1}{N} \left[\frac{1}{N_e^i} \int d^3r |\rho^i(\mathbf{r}) - \rho_{CCSD}^i(\mathbf{r})| \right], \quad (4.1)$$

where N_e is the number of electrons and N is the number of molecules. For conventional, non-empirical functionals (Tab. IV-4), we observe that the MAE decreases with increasing complexity of the functional: PBEh < MGGA < GGA. The empirical MN15L has the highest $\epsilon_{|n|}$, which suggests over-parametrization. The NN models that do not satisfy the HEG limits (NNx0, NNx2, NNx4) show that the errors in the density diminishes as more variables are added. However, NNx0 might be overfitted since its $\epsilon_{|n|}$ is larger than that of the PBE functional. For NNx2 and NNx4, the improvement observed for the MAE of the W4-11 and BH76 datasets compared to PBE are unlikely to be a result of overfitting, given that the densities are also improved. Overfitting is often observed for purely empirical functionals [63].

The models that satisfy the HEG limit (NNHx0, NNHx2 and NNHx4) all improve upon PBE, but the three of them share a similar $\epsilon_{|n|}$. The reason for this behavior might be the

Tab. IV-2. Calculated atomization energies in kcal/mol for the validation dataset of molecules with single and multiple bonds. MAE and ME are the mean average error and mean error, respectively. The calculations are done self-consistently for all functionals. The reference results are obtained from [60–62].

Molecule	PBE	PBEh	SCAN	MN15L	NNx0	NNx2	NNx4	NNHx0	NNHx2	NNHx4	Ref.
H ₂	105	104	107	110	109	108	107	108	106	108	109
LiH	53	52	55	62	49	50	53	50	52	51	58
CH ₄	420	417	418	417	419	417	417	417	416	418	419
NH ₃	299	292	292	293	291	290	289	294	289	291	297
H ₂ O	231	224	225	224	226	227	224	229	227	227	232
HF	140	135	135	136	137	139	137	139	139	138	141
Li ₂	20	19	18	24	15	15	19	18	19	18	24
LiF	135	128	130	131	122	127	132	126	134	130	139
Be ₂	10	6	8	3	6	6	7	7	7	7	3
MAE	3.3	5.7	4.9	3.7	6.1	5.6	5.0	4.7	4.5	4.7	0
ME	-1.2	-5.0	-3.7	-2.4	-5.3	-4.9	-4.2	-3.8	-3.6	-3.9	0
C ₂ H ₂	414	404	401	401	408	406	408	402	404	404	405
C ₂ H ₄	571	563	561	559	565	562	563	561	561	562	563
HCN	326	310	307	312	313	311	313	311	312	312	312
CO	269	255	254	261	258	260	261	261	264	261	259
N ₂	243	225	219	229	227	225	226	226	225	225	229
NO	172	153	151	152	159	158	157	160	159	158	153
O ₂	144	124	127	120	136	138	135	139	140	137	121
F ₂	53	35	37	39	41	45	43	48	49	46	39
P ₂	122	111	111	113	109	111	109	111	114	112	117
Cl ₂	65	59	58	61	60	62	60	64	64	63	58
MAE	12.4	2.6	4.3	1.8	4.2	4.5	4.0	5.6	5.7	4.7	0
ME	12.4	-1.9	-3.1	-0.7	1.9	2.1	1.9	2.7	3.5	2.5	0
Tot. MAE	8.1	4.1	4.6	2.7	5.1	5.0	4.5	5.2	5.1	4.7	0
Tot. ME	5.9	-3.3	-3.4	-1.5	-1.5	-1.2	-1.0	-0.4	0.1	-0.5	0

HEG constraint, reducing the flexibility of the NN. Without adding the electronic densities explicitly to the loss function, it is challenging to diminish $\epsilon_{|n|}$ in a systematic way, nonetheless, the NNs satisfying the HEG yield improved densities compared to PBE, as do NNx2 and NNx4.

4.4. Enhancement factor of a neural network functional

The plots for the X enhancement factor ($F_{X,\sigma}(\mathbf{r}) = \epsilon_{X,\sigma}(\mathbf{r})/\epsilon_{X,\sigma}^{HEG}(\mathbf{r})$) of NNx4 are provided in Fig. 4.2 and those of NNHx4 in Fig. 4.3 and they are compared to enhancement factors of NN(H)x0, PBE as well as B88 [64]. We chose the range of \mathcal{T}_σ values considered in the plots based on an analysis of chemical relevance for the N₂ molecule (see supplementary material). The range for the \mathcal{Q}_σ values is based on a previous study by our group [44]. For both \mathcal{Q}_σ

Model	W4-11		BH76	
	MAE	MAX	MAE	MAX
PBE	14.5	54	9.8	31
PBEh	4.2	27	4.2	14
SCAN	3.6	26	8.0	18
MN15L	4.1	24	2.4	10
NNx0	7.6	54	3.3	26
NNx2	7.3	34	4.2	26
NNx4	6.8	36	4.4	23
NNHx0	7.2	35	6.0	26
NNHx2	7.4	37	6.2	23
NNHx4	6.6	31	5.6	23

Tab. IV-3. Mean absolute errors (MAE) and maximal absolute error (MAX) in kcal/mol for an atomization energy dataset of 140 molecules (W4-11) and a barrier heights dataset of 76 systems (BH76). All the calculations are performed with self-consistent PBE orbitals.

Tab. IV-4. Errors in the self-consistent electron density relative to CCSD calculations (see Eq 4.1) for the dataset of molecules of Tab. IV-2

Method	$\epsilon_{ n }$
PBE	0.0100
PBEh	0.0062
SCAN	0.0065
MN15L	0.0135
NNx0	0.0116
NNx2	0.0091
NNx4	0.0084
NNHx0	0.0090
NNHx2	0.0096
NNHx4	0.0093

and \mathcal{T}_σ , the values corresponding to the HEG limit are also considered, if an approximation satisfies this limit, the enhancement factor is equal to 1 at $s_\sigma = 0$, such as shown for the PBE and B88 functional. For NNx4 and NNHx4, we can observe that the dependence on \mathcal{Q}_σ is larger than \mathcal{T}_σ . Despite the HEG limit not being a constraint for NNx4, Fig. 4.2 reveals that the NNx4 does not deviate much from the limit. The enhancement factors of NNx4 and NNHx4 are similar, indicating a certain stability of the NN representation with respect to changing constraints.

Overall, for NNx4 and NNHx4, the enhancement factors increase with increasing s_σ , there is a small decrease, however, for values of s around 1/2. The curvature \mathcal{Q}_σ has a large impact on the enhancement factor, for positive values of \mathcal{Q}_σ the X hole becomes more delocalized and the enhancement factor decreases. For negative \mathcal{Q}_σ , the X hole becomes more localized and the enhancement factor increases. This is the physically expected behavior

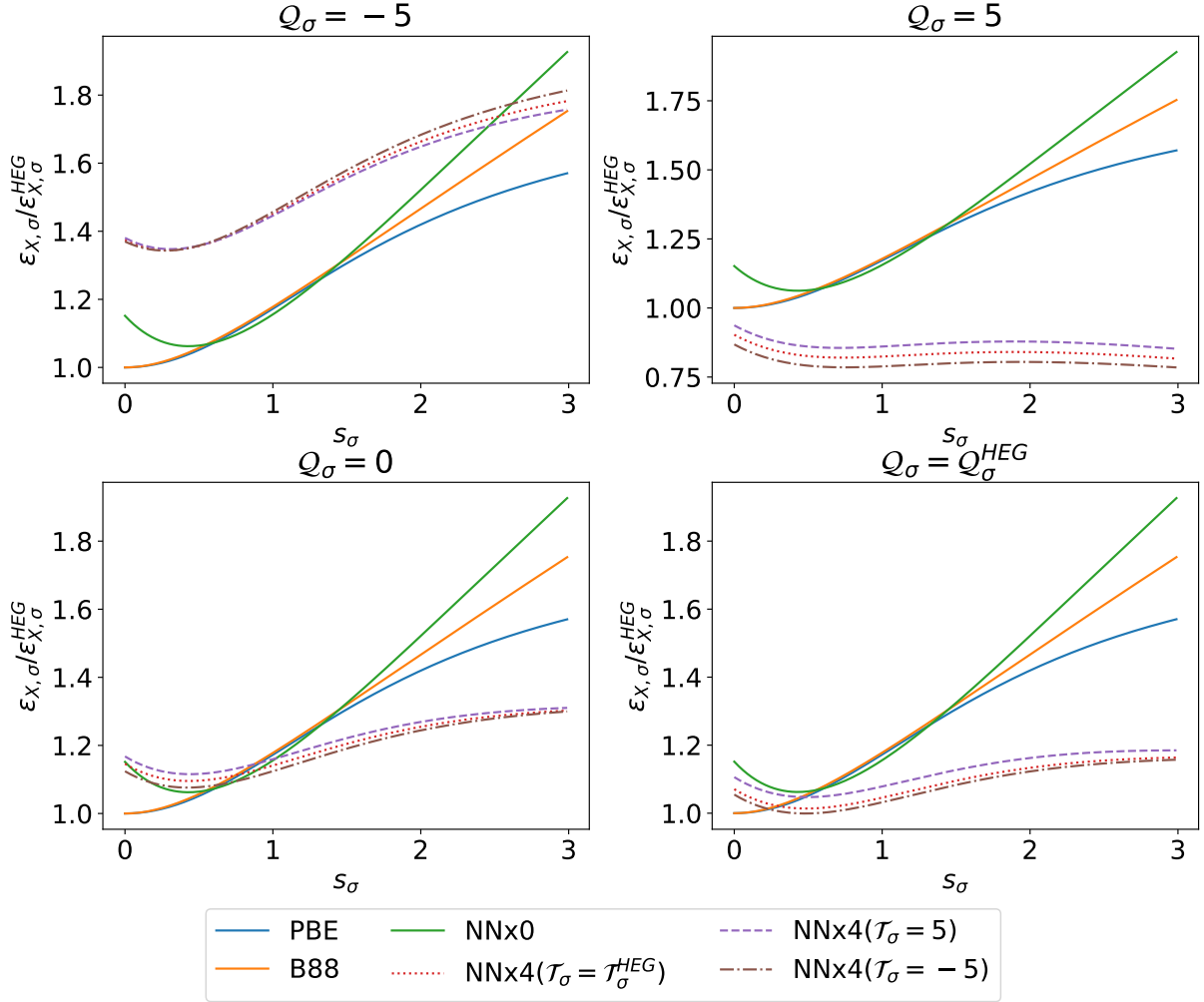


Fig. 4.2. The exchange enhancement factor ($\epsilon_{X, \sigma} / \epsilon_{X, \sigma}^{HEG}$) for PBE, B88, NNx0, and NNx4.

and the fourth-derivative \mathcal{T}_σ introduces a small variation in the enhancement factor, which is reassuring and it indicates a convergence of the enhancement factor with respect to increasing the order of the derivative used in the construction of functionals. For NNx4, we observe that the effect of the fourth derivative inverts direction as the curvature changes from negative to positive. For negative curvature, a negative \mathcal{T}_σ yields further enhancement, while for positive Q_σ the negative \mathcal{T}_σ results in reduction of the enhancement factor. In contrast, the \mathcal{T}_σ -dependence of NNx4 does not follow a simple pattern.

The enhancement factor of NNx0 shows a pronounced dip at $s_\sigma \approx 0.5$, which is absent in PBE and B88, it also increases for large s_σ -values compared to PBE and B88. The dip is likely to be the reason for the good performance of NNx0 for AE and it is also a sign of overfitting, biasing NNx0 towards molecules and strongly violating the HEG. For NNx4, the enforcement of the HEG limit generates an enhancement factor that oscillated around PBE and B88.

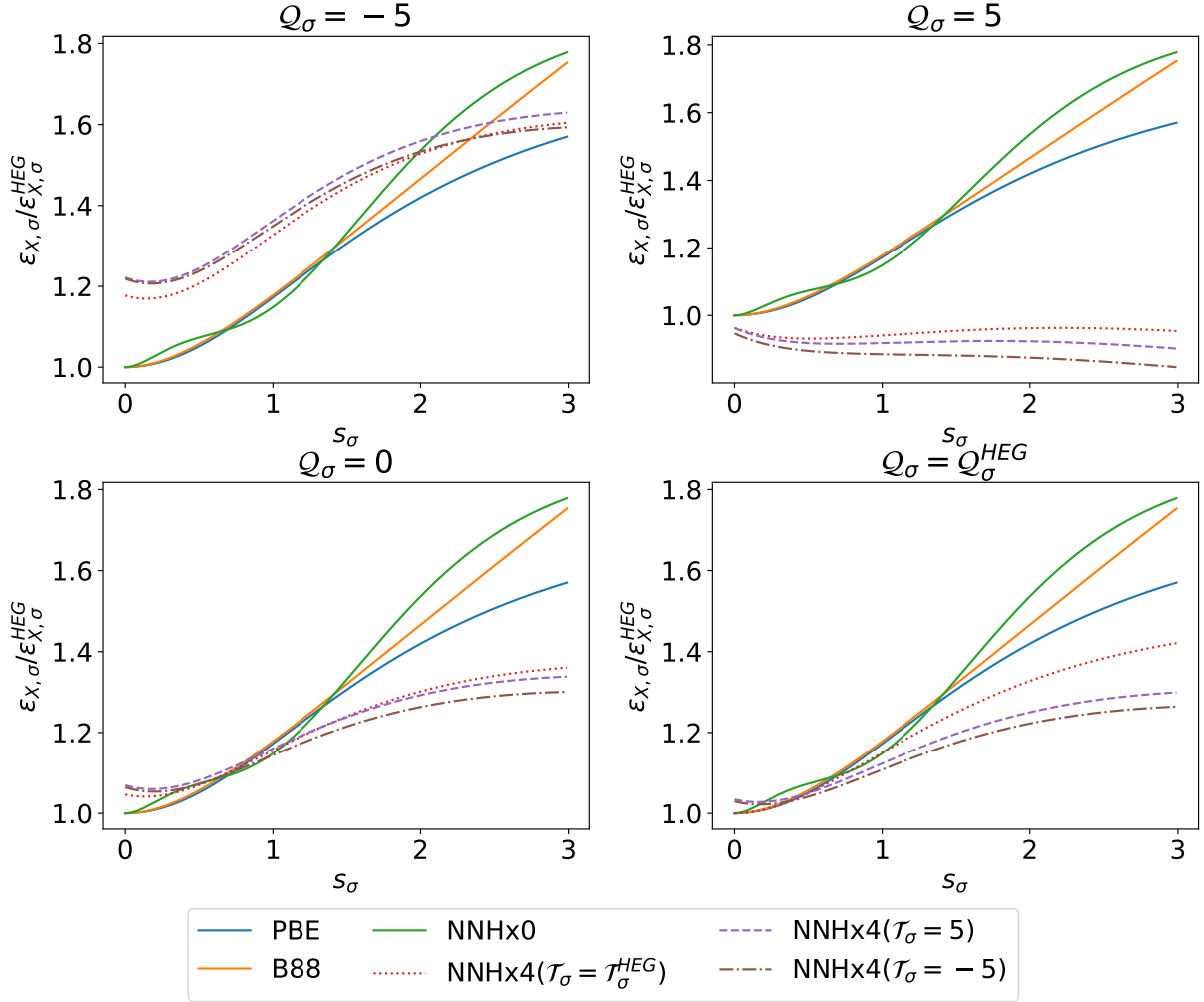


Fig. 4.3. The exchange enhancement factor ($\epsilon_{X,\sigma}/\epsilon_{X,\sigma}^{HEG}$) for PBE, B88, NNHx0, and NNHx4.

In the supplementary material, we also show the results obtained with a modified NN architecture for NNx4; despite using more neurons and a different activation function, the enhancement factor changes little, suggesting that the optimization procedure is not too sensitive to the NN's topology.

5. Conclusions

Often in the analysis of approximations for the XC energy, physical insight emerges, such as the correlation between the fourth-derivative of the X hole and non-localities in molecular X holes [19]. There is, however, no unique path to exploit such insight into the nature of XC to obtain improved approximations. To address this problem, here we develop a ML approach that can in principle be applied to any X functional to add a dependence on the fourth-order derivative T_σ . We consider the PBE X functional as an example and introduce

a T_σ -dependent enhancement factor for it. We don't expect the fourth-order expansion to be able to recover the full non-locality [65] of the X hole and we use ML to represent the X energy of PBEh which has a reduced degree of non-locality compared to the exact X hole; in this context, it is not desirable to cover more non-locality than is contained in the PBEh X energy since this then requires a corresponding non-local C energy functional.

We obtain a coherent, systematic approach in which the sequence of approximations NN(H)x0, NN(H)x2, NN(H)x4 shows that the fourth derivative is indeed useful to include non-localities that are not captured with s_σ and Q_σ . Employing only a few molecules to train a small network, we obtain an improvement on datasets of hundreds of systems for AE, BH, and the prediction of electron densities, even though neither BH nor electron densities are part of the training procedure. Since we represent a well defined degree of non-locality with the NNs, small training sets capturing this non-locality suffice to obtain a general improvement over PBE. The ML approach also has the advantage that the self-consistent implementation is simplified due to the automatic differentiation available in ML codes.

A next step in the development of XC functionals could be to design a T_σ -dependent C functional, this can be achieved with the correlation factor approach that we developed in our lab [47–51]. Starting from a given X functional, the corresponding X hole can be obtained and turned into an XC hole through the application of existing correlation factors.

6. Supplementary Material

The supplementary material contains technical details of the neural networks, of the electronic structure calculations, atomic exchange energy errors obtained for the various NN functionals, an analysis of the chemically relevant variable range of $\mathcal{T}_\sigma(\mathbf{r})$, an assessment of the impact of the activation function employed in the NN, as well as a study of the basis set dependence of the NN functionals. Furthermore, it contains the details of the self-consistent implementation of the NN functionals.

Acknowledgments

The financial support through *The Natural Sciences and Engineering Research Council of Canada (NSERC)* is gratefully acknowledged. EC acknowledges a scholarship from Fonds québécois de la recherche sur la nature et les technologies (FRQNT).

7. Data Availability Statement

The data that supports the findings of this study are available within the article and the supplementary material.

S1. Supplementary materials

S1.1. Technical details of the neural network and the electronic structure calculations

The neural networks (NN) are built with Pytorch [55] and Pytorch Lightning. To optimize the parameters of the NN, the limited-memory Broyden–Fletcher–Goldfarb–Shanno (LBFGS) optimization algorithm is employed. The inputs for the NN ($\mathbb{S}_\sigma, \mathbb{Q}_\sigma, \mathbb{T}_\sigma$) are standardized, thus their mean values are zero and the standard deviation is unity. The NN is considered converged when L_{REx} is lower than 2 kcal/mol to prevent overfitting. The NNs are optimized using floating numbers with 64 bit precision. For the calculation and to generate the training data, we use the Def2-tzvp [66, 67] basis sets with the PySCF software [68, 69]. The integration grid in PySCF is level=4, which corresponds to (radial shells, angular points per shell): (60,434), (90,590), (95,590) grids for atoms of period 1,2 and 3 respectively. The Python codes for the NN and its implementation within PySCF are available at the Gitlab webpage of our group (<https://gitlab.com/electronic-structure-udem>). None of the calculations reported, including the NN optimizations, require significant computational resources, they all run on a desktop computer.

S1.2. The mean absolute errors of atomic energies

The total atomic energies for atoms from H to Ar are calculated with the NN-based functionals and the reference functionals. The energies are compared to the accurate values of Refs.[70–72] and the MAE is obtained for each functionals. The results of Tab. IV.S1 show that the MAEs for our models are all close to PBEh. This is another confirmation that the NN representation results in sound functionals. The NNTx(0,2,4) functionals are obtained with a modified NN architecture, as described in Sec. S1.4.

S1.3. The fourth-order derivative of the exchange hole for the N_2 molecule

To provide insight into the behavior of the fourth-derivative and to examine the range of its values, we plot the values of the reduced fourth order derivative of the X hole \mathcal{T}_σ . In Fig. 4.S1, we observe that, at the nuclei, \mathcal{T}_σ oscillate considerably. For the chemically relevant valence region, the values are between -5 and 5. We use those values for the range of \mathcal{T}_σ to plot the enhancement factors in Figs. 4.2 and 4.3 of the main article.

Tab. IV.S1. The mean absolute errors (MAEs) of total energies of atoms ranging from H to Ar. For the models using neural networks, self-consistent PBE orbitals are used, while PBEh is calculated self-consistently. The reference energy are obtained from Refs.[70–72].

Model	MAE (Hartree)
PBE	0.089
PBEh	0.073
NNx0	0.076
NNx2	0.074
NNx4	0.072
NNHx0	0.074
NNHx2	0.074
NNHx4	0.075
NNTx0	0.074
NNTx2	0.069
NNTx4	0.073

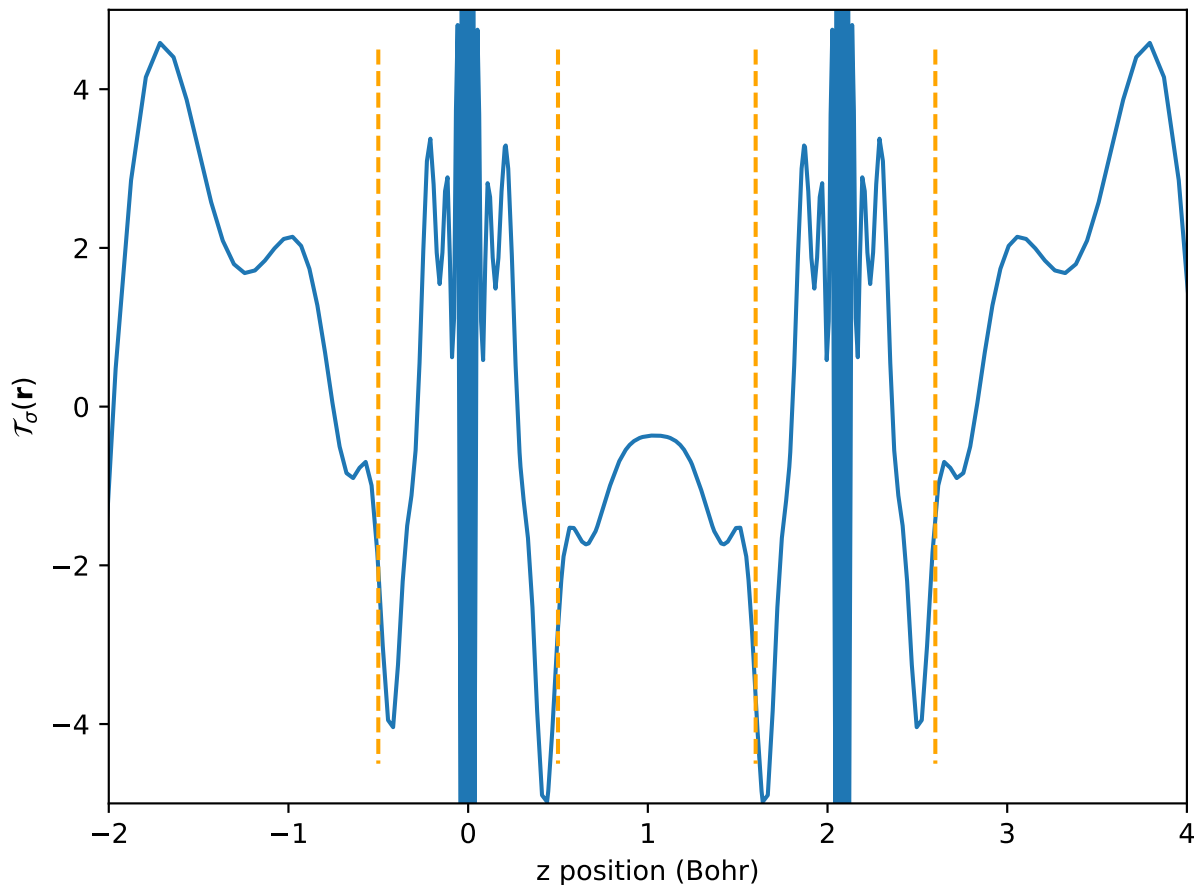


Fig. 4.S1. The values of the fourth derivative of the exchange hole $\mathcal{T}_\sigma(\mathbf{r})$ for the N_2 molecule along the z axis. The first atom is at the origin, while the second is at $(0,0,1.10)$ Angstrom. The orange dashed lines mark the transition from the core to the valence region.

Model	PBE	NNTx0	NNTx2	NNTx4
W4-11	16.1	6.1	6.3	5.0

Tab. IV.S2. Errors in the X contributions to atomization energies relative to PBEh, i.e. $\Delta E_X^{approx} - \Delta E^{PBEh}$. The mean absolute errors are calculated over the 140 molecules of the W4-11 data set [57] and is listed in kcal/mol. The calculation are done using self-consistent PBE orbitals

Model	W4-11		BH76	
	MAE	MAX	MAE	MAX
PBE	14.5	54	9.8	31
PBEh	4.2	27	4.2	14
SCAN	3.6	26	8.0	18
MN15L	4.1	24	2.4	10
NNTx0	7.6	51	3.3	26
NNTx2	7.3	32	6.2	23
NNTx4	6.4	38	5.3	23

Tab. IV.S3. Mean absolute errors (MAE) and maximal absolute error (MAX) in kcal/mol for the atomization energy dataset of 140 molecules (W4-11) and a barrier heights dataset of 76 systems (BH76). All the calculation are post-PBE.

S1.4. Impact of the choice of activation function in the neural network

To better understand the impact of the NN architecture on the resulting functionals we vary some of the technical aspects of the NN, the corresponding model is referred to as NNTx4. The modifications for the NN consists of changing the GELU activation function [55, 73] for the hyperbolic tangent (tanh) [55]. Furthermore, 16 neurons are used instead of 12. Table IV.S3 shows that the results for NNTx4 are slightly inferior to those obtained with GELU for barrier heights, but overall, the functionals constructed in this work are fairly robust with respect to changes of the NN’s architecture. Just as the models in the main text, in Tab. IV.S2, we observe that the exchange contribution to the atomization energies is systematically improved upon addition of the non-local variables. The addition of variables of increased non-locality also improves the atomization energies and barrier heights, albeit not in a systematic manner, as shown in Tab. IV.S3. Table IV.S4 shows that the electron densities are also improved, compared to PBE, by the functionals constructed with the modified NN architecture. The enhancement factor of NNTx4 is shown in Fig. 4.S2. Interestingly, its behavior is not significantly different from NNHx4 and NNx4 (see Figs. 2 and 3 in the main text). This indicates that our methodology is reproducible and that the NNs consistently include non-local information generalized from the input data.

Tab. IV.S4. Error on the calculated electron density compared to CCSD calculations (see equation 4.1 in the main text) for the dataset of molecules of Tab. IV-2 of the main text.

Model	$\epsilon_{ n }$
PBE	0.0100
PBE0	0.0062
SCAN	0.0065
MN15L	0.0135
NNTx0	0.0115
NNTx2	0.0091
NNTx4	0.0084

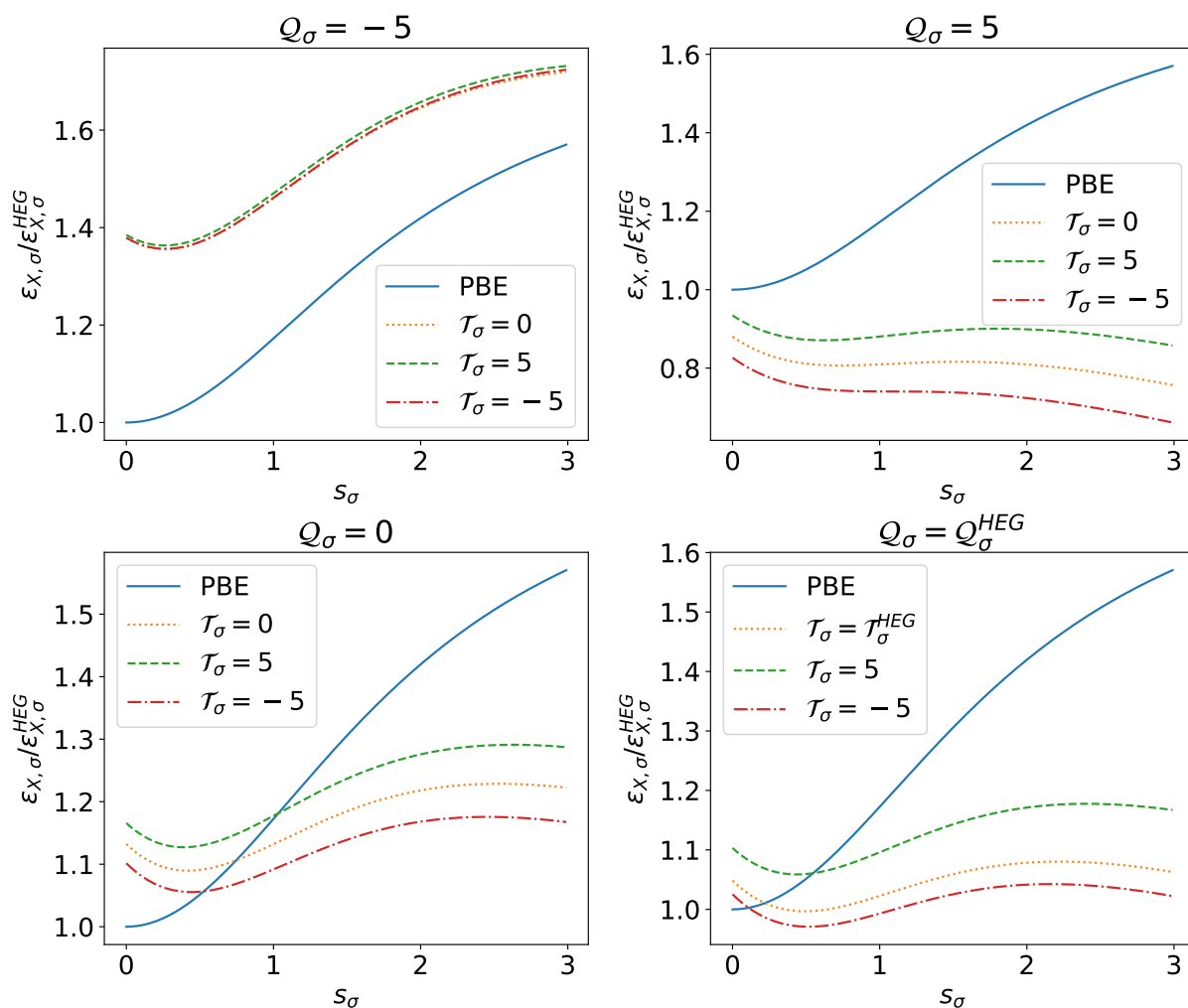


Fig. 4.S2. The exchange enhancement factor for PBE and the NNTx4 model for various Q_σ and \mathcal{T}_σ .

Tab. IV.S5. The mean absolute errors (MAE) for total energies (E^{tot}) of atoms (H to Ar) and for the atomization energies (AE) of the 4 dataset [56, 57], obtained with different gaussian basis sets (def2-SVP [66], def2-TZVP [67], and cc-PVTZ [74–78]).

Basis	MAE for E^{tot} (Ha)			MAE for AEs (kcal/mol)		
	def2-SVP	def2-TZVP	cc-pVTZ	def2-SVP	def2-TZVP	cc-pVTZ
PBE	0.172	0.089	0.089	14.7	14.5	13.5
PBEh	0.154	0.073	0.072	6.5	4.2	4.2
NNx4	0.163	0.072	0.069	8.7	6.8	6.8

S1.5. Basis set dependence of \mathcal{T}_σ -dependent functionals

In a previous study of the fourth-order term of the expansion of the X hole [19], we observed a strong basis set dependence of this quantity, particularly in the core region of atoms. Therefore, it is pertinent to analyze the basis set dependence of NNx4. The basis sets considered are def2-SVP [66] (e.g. 3s2p1d contracted functions for N), def2-TZVP [67] (e.g. 5s3p2d1f contracted functions for N), and cc-PVTZ [74–78] (e.g. 4s3p2d1f for N). For the total energy and atomization energy calculations (Tab. IV.S5), NNx4 shows a basis set dependence no more pronounced than PBEh. As mentioned in the article, the log transformation of the \mathcal{T}_σ variable suppresses the large, oscillating \mathcal{T}_σ -values in the core region. Also, the NN training process suppresses the somewhat erratic behavior of \mathcal{T}_σ in the core region and these factors result in a basis set dependence comparable to conventional functionals.

S1.6. Dissociation of H_2+

In Fig. 4.S3 we consider the dissociation of H_2+ , it reveals that, compared to PBE, the NNx4 follows PBEh more closely than PBE does and thus it includes an additional degree of non-locality relative to PBE.

S1.7. Self-consistent implementation of the NN functionals

The self-consistent implementation is performed within the generalized Kohn-Sham scheme in which the energy is made stationary with respect to the density matrix elements, as outlined in Ref.[79]. The electronic energy as a function of the density matrix P_{uv} in the atomic orbital basis (χ_i) is,

$$\begin{aligned}
 E = \sum_{uv} P_{uv} \int d\mathbf{r} \chi_u(\mathbf{r}) \left(-\frac{1}{2} \nabla^2 + \sum_N \frac{Z_N}{r_N} \right) \chi_v(\mathbf{r}) \\
 + \frac{1}{2} \sum_{uv\lambda\sigma} P_{uv} P_{\lambda\sigma} \int d\mathbf{r} \int d\mathbf{r}' \chi_u(\mathbf{r}) \chi_v(\mathbf{r}) \frac{1}{r_{12}} \chi_\lambda(\mathbf{r}') \chi_\sigma(\mathbf{r}') + \int d\mathbf{r} \rho(\mathbf{r}) \epsilon_{xc}(\mathbf{r}). \quad (\text{S1})
 \end{aligned}$$

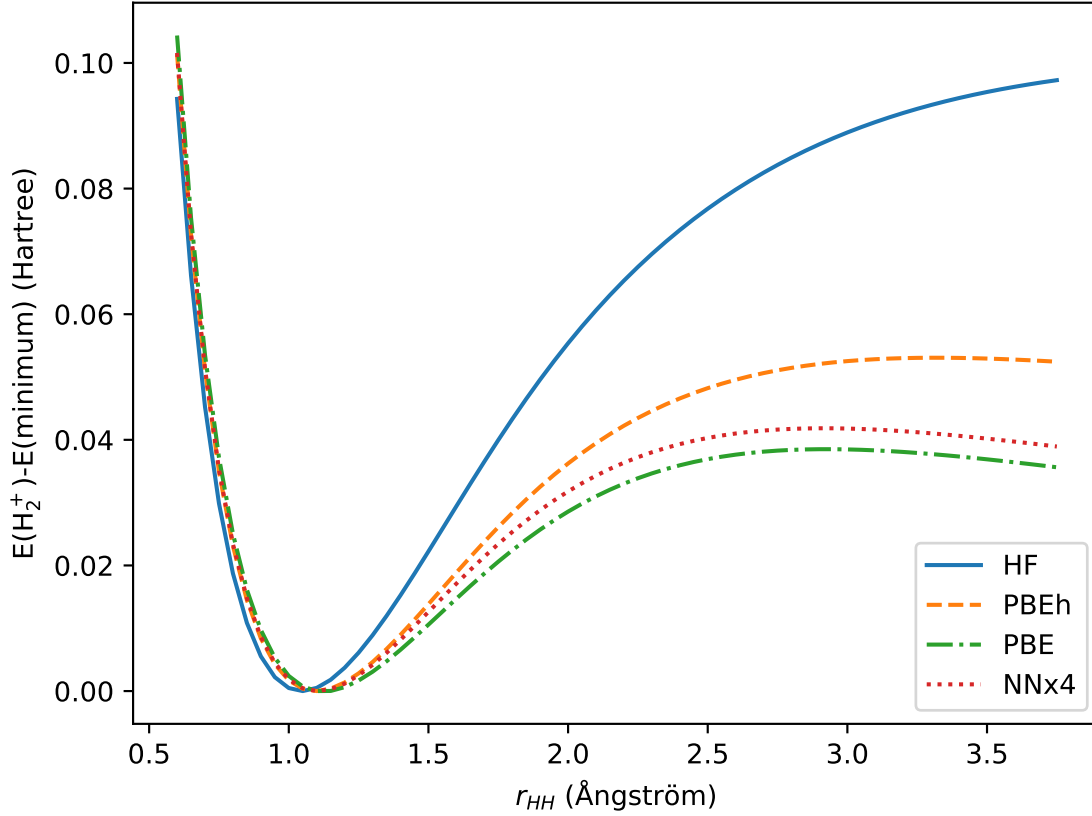


Fig. 4.S3. Self-consistent calculation of the H_2^+ dissociation curve with NNx4 compared to the exact HF, PBEh, and PBE. The curves have been shifted such that the minima coincide.

The equations for the self-consistent implementation of the first two terms can be found in Ref.[79], in this section we will focus on the last term,

$$E_{XC} = E_C^{PBE} + E_X^\sigma + E_X^{\sigma'}. \quad (\text{S2})$$

The derivative of E_C^{PBE} with respect to the density matrix is already implemented in PySCF, whereas the derivative of the X energy with respect to the density matrix has to be calculated,

$$\begin{aligned} \frac{\partial E_{X,\sigma}}{\partial P_{uv,\sigma}} &= \frac{\partial \int d^3r f(\mathbb{S}_\sigma, \mathbb{Q}_\sigma, \mathbb{T}_\sigma) e_{X,\sigma}^{PBE}(\mathbf{r})}{\partial P_{uv,\sigma}} \\ &= \int d^3r \left(f(\mathbb{S}_\sigma, \mathbb{Q}_\sigma, \mathbb{T}_\sigma) \frac{\partial e_{X,\sigma}^{PBE}(\mathbf{r})}{\partial P_{uv,\sigma}} + e_{X,\sigma}^{PBE} \frac{\partial f(\mathbb{S}_\sigma, \mathbb{Q}_\sigma, \mathbb{T}_\sigma)(\mathbf{r})}{\partial P_{uv,\sigma}} \right). \end{aligned} \quad (\text{S3})$$

$\frac{\partial e^{\mathcal{X}_{\sigma}}(\mathbf{r})}{\partial P_{uv,\sigma}}$ in this equation is known and won't be discussed further. Now we calculate $\frac{\partial f(\mathcal{S}_{\sigma}, \mathcal{Q}_{\sigma}, \mathcal{T}_{\sigma})}{\partial P_{uv,\sigma}}$,

$$\frac{\partial f(s_{\sigma}, \mathcal{Q}_{\sigma}, \mathcal{T}_{\sigma})}{\partial P_{uv,\sigma}} = \frac{\partial f}{\partial \mathcal{S}_{\sigma}} \frac{\partial \mathcal{S}_{\sigma}}{\partial s_{\sigma}} \frac{\partial s_{\sigma}}{\partial P_{uv,\sigma}} + \frac{\partial f}{\partial \mathcal{Q}_{\sigma}} \frac{\partial \mathcal{Q}_{\sigma}}{\partial \mathcal{Q}_{\sigma}} \frac{\partial \mathcal{Q}_{\sigma}}{\partial P_{uv,\sigma}} + \frac{\partial f}{\partial \mathcal{T}_{\sigma}} \frac{\partial \mathcal{T}_{\sigma}}{\partial \mathcal{T}_{\sigma}} \frac{\partial \mathcal{T}_{\sigma}}{\partial P_{uv,\sigma}}. \quad (\text{S4})$$

Using automatic differentiation, we obtain the derivatives of the NN representations of the enhancement factors. In addition, we need the derivatives of the various input variables with respect to the density matrix, i.e., $\frac{\partial s_{\sigma}}{\partial P_{uv,\sigma}}$, $\frac{\partial \mathcal{Q}_{\sigma}}{\partial P_{uv,\sigma}}$, $\frac{\partial \mathcal{T}_{\sigma}}{\partial P_{uv,\sigma}}$. In the following sections, we calculate these three derivatives. For simplicity of notation, we drop σ and \mathbf{r} in the equations.

S1.7.1. Derivative of s with respect to P_{uv} .

First we define ρ and $\nabla\rho$ as a function of the density matrix,

$$\rho = \sum_{uv} P_{uv} \chi_u \chi_v, \quad (\text{S5})$$

where χ_i are the atomic orbitals. The gradient of the density is

$$\nabla\rho = \sum_{uv} P_{uv} \nabla(\chi_u \chi_v). \quad (\text{S6})$$

The derivative of the reduced gradient s with respect to the density matrix is,

$$\frac{\partial s}{\partial P_{uv}} = \frac{k_F \rho \frac{\nabla\rho}{|\nabla\rho|} \cdot \frac{\partial \nabla\rho}{\partial P_{uv}} - |\nabla\rho| \frac{\partial k_F \rho}{\partial P_{uv}}}{2k_F^2 \rho^2}, \quad (\text{S7})$$

with

$$\frac{\partial \nabla\rho}{\partial P_{uv}} = 2\chi_u \nabla\chi_v, \quad (\text{S8})$$

$$\frac{\partial k_F \rho}{\partial P_{uv}} = \frac{4}{3} k_F \frac{\partial \rho}{\partial P_{uv}}, \quad (\text{S9})$$

and

$$\frac{\partial \rho}{\partial P_{uv}} = \chi_u \chi_v. \quad (\text{S10})$$

S1.7.2. Derivative of Q with respect to P_{uv} .

We define τ and $\nabla^2\rho$ as functions of the density matrix

$$\tau = \sum_{uv} P_{uv} \nabla\chi_u \nabla\chi_v \quad (\text{S11})$$

$$\nabla^2\rho = \sum_{uv} P_{uv} \nabla^2(\chi_u \chi_v). \quad (\text{S12})$$

Just as for s , we change from the reduced curvature to the curvature

$$\frac{\partial Q/k_F^2 \rho}{\partial P_{uv}} = \frac{k_F^2 \rho \frac{\partial Q}{\partial P_{uv}} - Q \frac{\partial k_F^2 \rho}{\partial P_{uv}}}{k_F^4 \rho^2}, \quad (\text{S13})$$

with

$$\frac{\partial Q}{\partial P_{uv}} = \frac{1}{6} \left(\frac{\partial \nabla^2 \rho}{\partial P_{uv}} - 2 \frac{\partial \tau}{\partial P_{uv}} + \frac{1}{2} \frac{2\rho \nabla \rho \cdot \frac{\partial \nabla \rho}{\partial P_{uv}} - (\nabla \rho)^2 \frac{\partial \rho}{\partial P_{uv}}}{\rho^2} \right). \quad (\text{S14})$$

The derivative of ρ and $\nabla \rho$ with respect to the density matrix is already given in the previous section and

$$\frac{\partial \nabla^2 \rho}{\partial P_{uv}} = \chi_u \nabla^2 \chi_v + 2 \frac{\partial \tau}{\partial P_{uv}}, \quad (\text{S15})$$

with

$$\frac{\partial \tau}{\partial P_{uv}} = \nabla \chi_u \nabla \chi_v. \quad (\text{S16})$$

S1.7.3. Derivative of T with respect to P_{uv} .

We focus on $\frac{\partial T}{\partial P_{uv}}$. T is the sum of four terms, which we differentiate individually,

$$T(\mathbf{r}) = \frac{1}{5! \rho} [T1 + T2 + T3 + T4] \quad (\text{S17})$$

In atomic orbital basis, we have

$$T1 = 2\rho \left(\sum_{uv} P_{uv} \chi_u \nabla^4 \chi_v \right) \quad (\text{S18})$$

$$\frac{\partial T1}{\partial P_{uv}} = 2 \frac{\partial \rho}{\partial P_{uv}} \sum_{uv} P_{uv} \chi_u \nabla^4 \chi_v + 2\rho (\chi_u \nabla^4 \chi_v) \quad (\text{S19})$$

For T2,

$$T2 = 4 \nabla \rho \cdot \left(\sum_{uv} P_{uv} \chi_u \nabla^3 \chi_v \right) \quad (\text{S20})$$

$$\frac{\partial T2}{\partial P_{uv}} = 4 \frac{\partial \nabla \rho}{\partial P_{uv}} \cdot \left(\sum_{uv} P_{uv} \chi_u \nabla^3 \chi_v \right) + 4 \nabla \rho \cdot (\chi_u \nabla^3 \chi_v) \quad (\text{S21})$$

For T3,

$$T3 = \frac{1}{2} (\nabla^2 \rho - 2\tau)^2 \quad (\text{S22})$$

$$\frac{\partial T3}{\partial P_{uv}} = (\nabla^2 \rho - 2\tau) \left(\frac{\partial \nabla^2 \rho}{\partial P_{uv}} - 2 \frac{\partial \tau}{\partial P_{uv}} \right) \quad (\text{S23})$$

For T4,

$$T4 = 4 \sum_{uv} P_{uv} \chi_u \left(\frac{\partial^2 \chi_v}{\partial x^2}, \frac{\partial^2 \chi_v}{\partial y^2}, \frac{\partial^2 \chi_v}{\partial z^2}, \sqrt{2} \frac{\partial^2 \chi_v}{\partial x \partial y}, \sqrt{2} \frac{\partial^2 \chi_v}{\partial x \partial z}, \sqrt{2} \frac{\partial^2 \chi_v}{\partial y \partial z} \right). \quad (\text{S24})$$

$$\sum_{ij} P_{ij} \chi_i \left(\frac{\partial^2 \chi_j}{\partial x^2}, \frac{\partial^2 \chi_j}{\partial y^2}, \frac{\partial^2 \chi_j}{\partial z^2}, \sqrt{2} \frac{\partial^2 \chi_j}{\partial x \partial y}, \sqrt{2} \frac{\partial^2 \chi_j}{\partial x \partial z}, \sqrt{2} \frac{\partial^2 \chi_j}{\partial y \partial z} \right) \quad (\text{S25})$$

$$\frac{\partial T4}{\partial P_{uv}} = 8 \sum_{uv} P_{uv} \chi_u \left(\frac{\partial^2 \chi_v}{\partial x^2}, \frac{\partial^2 \chi_v}{\partial y^2}, \frac{\partial^2 \chi_v}{\partial z^2}, \sqrt{2} \frac{\partial^2 \chi_v}{\partial x \partial y}, \sqrt{2} \frac{\partial^2 \chi_v}{\partial x \partial z}, \sqrt{2} \frac{\partial^2 \chi_v}{\partial y \partial z} \right). \quad (\text{S26})$$

$$\chi_i \left(\frac{\partial^2 \chi_j}{\partial x^2}, \frac{\partial^2 \chi_j}{\partial y^2}, \frac{\partial^2 \chi_j}{\partial z^2}, \sqrt{2} \frac{\partial^2 \chi_j}{\partial x \partial y}, \sqrt{2} \frac{\partial^2 \chi_j}{\partial x \partial z}, \sqrt{2} \frac{\partial^2 \chi_j}{\partial y \partial z} \right). \quad (\text{S27})$$

Thus,

$$\frac{\partial T}{\partial P_{uv}} = \frac{1}{120\rho} \left(\frac{\partial T_1}{\partial P_{uv}} + \frac{\partial T_2}{\partial P_{uv}} + \frac{\partial T_3}{\partial P_{uv}} + \frac{\partial T_4}{\partial P_{uv}} - \frac{\partial \rho}{\partial P_{uv}} \left(\frac{T_1 + T_2 + T_3 + T_4}{\rho} \right) \right). \quad (\text{S28})$$

References

- [1] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
- [2] W. Kohn and L. J. Sham, Physical Review **140**, A1133 (1965).
- [3] R. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules (New York, 1989)*. (Oxford University Press, New York, 1989).
- [4] R. Dreizler and E. Gross, *Density-Functional Theory (Springer, Berlin, 1990)* (Springer, 1990).
- [5] A. D. Becke, The Journal of Chemical Physics **140**, 18A301 (2014).
- [6] B. G. Janesko, Chem. Soc. Rev. **50**, 8470 (2021).
- [7] J. P. Perdew and K. Schmidt, AIP Conference Proceedings **577**, 1 (2001).
- [8] J. P. Perdew, K. Burke, and M. Ernzerhof, Physical Review Letters **77**, 3865 (1996).
- [9] J. Sun, A. Ruzsinszky, and J. P. Perdew, Phys. Rev. Lett. **115**, 036402 (2015).
- [10] A. D. Becke, The Journal of Chemical Physics **98**, 5648 (1993).
- [11] C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).
- [12] J. P. Perdew, M. Ernzerhof, and K. Burke, The Journal of Chemical Physics **105**, 9982 (1996).
- [13] C. Adamo and V. Barone, The Journal of Chemical Physics **110**, 6158 (1999).
- [14] A. J. Cohen, P. Mori-Sánchez, and W. Yang, Science **321**, 792 (2008).
- [15] K. Bystrom and B. Kozinsky, Journal of Chemical Theory and Computation , acs.jctc.1c00904 (2022).
- [16] M. Ernzerhof and J. P. Perdew, The Journal of Chemical Physics **109**, 3313 (1998).
- [17] A. D. Becke, International Journal of Quantum Chemistry **23**, 1915 (1983).
- [18] A. D. Becke and M. R. Roussel, Physical Review A **39**, 3761 (1989).
- [19] R. Wang, Y. Zhou, and M. Ernzerhof, Physical Review A **96**, 022502 (2017).
- [20] B. Kalita, L. Li, R. J. McCarty, and K. Burke, Accounts of Chemical Research **54**, 818 (2021).
- [21] S. Manzhos, Machine Learning: Science and Technology **1**, 013002 (2020).
- [22] J. A. Keith, V. Vassilev-Galindo, B. Cheng, S. Chmiela, M. Gastegger, K.-R. Müller, and A. Tkatchenko, arXiv:2102.06321 [physics] (2021).
- [23] J. Westermayr, M. Gastegger, K. T. Schütt, and R. J. Maurer, The Journal of Chemical Physics **154**, 230903 (2021).
- [24] L. Fiedler, K. Shah, M. Bussmann, and A. Cangi, Physical Review Materials **6**, 040301 (2022).
- [25] R. Nagai and R. Akashi, Development of Exchange-Correlation Functionals Assisted by Machine Learning (2022), arXiv:2206.15370 .
- [26] K. Ryczko, D. A. Strubbe, and I. Tamblyn, Physical Review A **100**, 022512 (2019).

- [27] M. Bogojeski, L. Vogt-Maranto, M. E. Tuckerman, K.-R. Müller, and K. Burke, *Nature Communications* **11**, 5223 (2020).
- [28] F. Brockherde, L. Vogt, L. Li, M. E. Tuckerman, K. Burke, and K.-R. Müller, *Nature Communications* **8**, 872 (2017).
- [29] J. C. Snyder, M. Rupp, K. Hansen, K.-R. Müller, and K. Burke, *Physical Review Letters* **108**, 253002 (2012).
- [30] J. Hollingsworth, T. E. Baker, and K. Burke, *The Journal of Chemical Physics* **148**, 241743 (2018).
- [31] J. Kirkpatrick, B. McMorrow, D. H. P. Turban, A. L. Gaunt, J. S. Spencer, A. G. D. G. Matthews, A. Obika, L. Thiry, M. Fortunato, D. Pfau, L. R. Castellanos, S. Petersen, A. W. R. Nelson, P. Kohli, P. Mori-Sánchez, D. Hassabis, and A. J. Cohen, *Science* **374**, 1385 (2021).
- [32] R. Nagai, R. Akashi, S. Sasaki, and S. Tsuneyuki, *The Journal of Chemical Physics* **148**, 241737 (2018).
- [33] R. Nagai, R. Akashi, and O. Sugino, *npj Computational Materials* **6**, 43 (2020).
- [34] L. Li, S. Hoyer, R. Pederson, R. Sun, E. D. Cubuk, P. Riley, and K. Burke, *Physical Review Letters* **126**, 036401 (2021).
- [35] B. Kalita, R. Pederson, J. Chen, L. Li, and K. Burke, *The Journal of Physical Chemistry Letters* , 2540 (2022).
- [36] M. F. Kasim and S. M. Vinko, *Phys. Rev. Lett.* **127**, 126403 (2021).
- [37] S. Dick and M. Fernandez-Serra, *Nature Communications* **11**, 3509 (2020).
- [38] S. Dick and M. Fernandez-Serra, *Physical Review B* **104**, L161109 (2021).
- [39] J. Schmidt, C. L. Benavides-Riveros, and M. A. L. Marques, *The Journal of Physical Chemistry Letters* **10**, 6425 (2019).
- [40] J. Gedeon, J. Schmidt, M. J. P. Hodgson, J. Wetherell, C. L. Benavides-Riveros, and M. A. L. Marques, *arXiv:2106.16075 [cond-mat, physics:physics, physics:quant-ph]* (2021).
- [41] A. Ryabov, I. Akhatov, and P. Zhilyaev, *Scientific Reports* **10**, 8000 (2020).
- [42] R. Nagai, R. Akashi, and O. Sugino, *Phys. Rev. Research* **4**, 013106 (2022).
- [43] K. Trepte and J. Voss, *arXiv:2201.11106 [cond-mat, physics:physics]* (2022).
- [44] E. Cuierrier, P.-O. Roy, and M. Ernzerhof, *The Journal of Chemical Physics* **155**, 174121 (2021).
- [45] X. Lei and A. J. Medford, *Phys. Rev. Materials* **3**, 063801 (2019).
- [46] B. G. Janesko, *The Journal of Chemical Physics* **133**, 104103 (2010).
- [47] J. Přecechtělová, H. Bahmann, M. Kaupp, and M. Ernzerhof, *The Journal of Chemical Physics* **141**, 111102 (2014).
- [48] J. Přecechtělová, H. Bahmann, M. Kaupp, and M. Ernzerhof, *The Journal of Chemical Physics* **143**, 144102 (2015).

- [49] R. Wang, Y. Zhou, and M. Ernzerhof, *The Journal of Chemical Physics* **150**, 084107 (2019).
- [50] R. Wang, Y. Zhou, and M. Ernzerhof, *The Journal of Chemical Physics* **151**, 194102 (2019).
- [51] P.-O. Roy, E. Cuierrier, and M. Ernzerhof, *The Journal of Chemical Physics* **152**, 211101 (2020).
- [52] M. Levy and J. P. Perdew, *Phys. Rev. A* **32**, 2010 (1985).
- [53] A. LeNail, *Journal of Open Source Software* **4**, 747 (2019).
- [54] H. S. Yu, X. He, and D. G. Truhlar, *Journal of Chemical Theory and Computation* **12**, 1280 (2016).
- [55] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimeshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, arXiv:1912.01703 [cs, stat] (2019).
- [56] A. Karton, S. Daon, and J. M. Martin, *Chemical Physics Letters* **510**, 165 (2011).
- [57] L. Goerigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi, and S. Grimme, *Phys. Chem. Chem. Phys.* **19**, 32184 (2017).
- [58] M. Ernzerhof, J. P. Perdew, and K. Burke, *International Journal of Quantum Chemistry* **64**, 285 (1997).
- [59] L. Goerigk and S. Grimme, *Physical Chemistry Chemical Physics* **13**, 6670 (2011).
- [60] J. A. Pople, M. Head-Gordon, D. J. Fox, K. Raghavachari, and L. A. Curtiss, *The Journal of Chemical Physics* **90**, 5622 (1989).
- [61] L. A. Curtiss, C. Jones, G. W. Trucks, K. Raghavachari, and J. A. Pople, *The Journal of Chemical Physics* **93**, 2537 (1990).
- [62] V. E. Bondybey and J. H. English, *The Journal of Chemical Physics* **80**, 568 (1984).
- [63] M. G. Medvedev, I. S. Bushmarinov, J. Sun, J. P. Perdew, and K. A. Lyssenko, *Science* **355**, 49 (2017).
- [64] A. D. Becke, *Physical Review A* **38**, 3098 (1988).
- [65] A. D. Becke, *The Journal of Chemical Physics* **119**, 2972 (2003).
- [66] F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.* **7**, 3297 (2005).
- [67] F. Weigend, *Phys. Chem. Chem. Phys.* **8**, 1057 (2006).
- [68] Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. McClain, E. R. Sayfutyarova, S. Sharma, S. Wouters, and G. K.-L. Chan, arXiv:1701.08223 [physics] (2017).
- [69] Q. Sun, X. Zhang, S. Banerjee, P. Bao, M. Barbry, N. S. Blunt, N. A. Bogdanov, G. H. Booth, J. Chen, Z.-H. Cui, J. J. Eriksen, Y. Gao, S. Guo, J. Hermann, M. R. Hermes, K. Koh, P. Koval, S. Lehtola, Z. Li, J. Liu, N. Mardirossian, J. D. McClain, M. Motta, B. Mussard, H. Q. Pham, A. Pulkin, W. Purwanto, P. J. Robinson, E. Ronca,

- E. Sayfutyarova, M. Scheurer, H. F. Schurkus, J. E. T. Smith, C. Sun, S.-N. Sun, S. Upadhyay, L. K. Wagner, X. Wang, A. White, J. D. Whitfield, M. J. Williamson, S. Wouters, J. Yang, J. M. Yu, T. Zhu, T. C. Berkelbach, S. Sharma, A. Sokolov, and G. K.-L. Chan, *The Journal of Chemical Physics* **153**, 024109 (2020).
- [70] C. L. Pekeris, *Phys. Rev.* **126**, 1470 (1962).
- [71] C. Schwartz, *International Journal of Modern Physics E* **15**, 877 (2006).
- [72] S. J. Chakravorty, S. R. Gwaltney, E. R. Davidson, F. A. Parpia, and C. F. P. Fischer, *Phys. Rev. A* **47**, 3649 (1993).
- [73] D. Hendrycks and K. Gimpel, *Gaussian Error Linear Units (GELUs)* (2020), arXiv:1606.08415 [cs.LG] .
- [74] T. H. Dunning, *The Journal of Chemical Physics* **90**, 1007 (1989).
- [75] R. A. Kendall, T. H. Dunning, and R. J. Harrison, *The Journal of Chemical Physics* **96**, 6796 (1992).
- [76] B. P. Prascher, D. E. Woon, K. A. Peterson, T. H. Dunning, and A. K. Wilson, *Theoretical Chemistry Accounts* **128**, 69 (2011).
- [77] D. E. Woon and T. H. Dunning, *The Journal of Chemical Physics* **98**, 1358 (1993).
- [78] D. E. Woon and T. H. Dunning, *The Journal of Chemical Physics* **100**, 2975 (1994).
- [79] S. Lehtola, F. Blockhuys, and C. Van Alsenoy, *Molecules* **25** (2020).

Conclusions

La théorie de la fonctionnelle de la densité (DFT) moderne a été développée durant les années 1960 et a mené au prix de Nobel de chimie de 1998 à Walter Kohn. La DFT est devenue la méthode de choix en chimie computationnelle pour prédire les propriétés électroniques des molécules. Par exemple, la DFT permet de mieux comprendre les réactions chimiques en chimie organique ou de surface. L'industrie pharmaceutique s'en sert pour trouver de nouveaux médicaments, tout comme l'industrie des matériaux qui cherche constamment à améliorer la performance des matériaux. Depuis presque 60 ans, la DFT est un domaine de recherche qui évolue constamment grâce aux nouvelles technologies et aux nouveaux algorithmes.

Le premier théorème de Hohenberg-Kohn prouve que la densité électronique contient toute l'information nécessaire pour calculer l'énergie électronique d'une molécule. Ensuite, Kohn et Sham (KS) ont développé une approche pour calculer cette énergie en pratique. L'approche KS est une alternative à la résolution de l'équation de Schrödinger, mais seulement un de ses termes est inconnu et doit être approximé en pratique: la fonctionnelle d'échange-corrélation (XC). L'histoire du développement des approximations en DFT est connectée au développement des nouvelles technologies, les premières approximations étaient plus rudimentaires, car les ordinateurs de l'époque ne permettaient pas d'effectuer des calculs aussi complexes qu'aujourd'hui. L'approximation de la densité locale (LDA) ne contient qu'un seul ingrédient qui est la densité électronique. La LDA n'est pas une approximation assez fine pour donner des résultats satisfaisants pour prédire la structure électronique des molécules, mais au fil des années, d'autres variables se sont ajoutées comme le gradient de la densité ou la densité d'énergie d'échange exact par particule. Ce coût computationnel plus important permet cependant d'avoir une meilleure précision pour des systèmes plus complexes comme les états de transition.

Un autre développement technologique important est l'apprentissage automatique (ML), qui trouve des relations dans des données massives, une tâche presque impossible à accomplir à la main par un humain. La DFT voit même des compagnies comme DeepMind (entreprise d'intelligence artificielle appartenant à Google) utiliser le ML pour construire une

approximation à l'énergie XC en DFT. L'exactitude des prédictions provenant des nouvelles fonctionnelles en DFT augmente, mais les fonctionnelles sont de plus en plus compliquées.

Dans cette thèse, nous pouvons aussi apercevoir une évolution similaire dans les projets où de nouvelles technologies sont utilisées pour régler des problèmes complexes. Le premier projet a tenté de construire des approximations non locales à partir d'équations mathématiques simples, ce qui a donné des résultats inférieurs à ceux des fonctionnelles hybrides pour la prédiction de propriétés thermochimiques de molécules. Donc, les énergies d'atomisation calculées ont des erreurs plus élevées que les fonctionnelles courantes en DFT. Le projet a quand même été utile pour mieux comprendre les modèles non locaux en DFT. Depuis environ 1970, des méthodes utilisant la densité non locale sont développées, malgré de nouvelles approches comme le facteur de corrélation, les prédictions de nos modèles non locales pour les molécules sont inférieures aux fonctionnelles XC courantes en DFT. La densité non locale montre des oscillations importantes, donc une nouvelle approche flexible est nécessaire.

L'échec de la construction d'une fonctionnelle non locale prédisant l'énergie électronique avec une exactitude satisfaisante a donc inspiré l'utilisation d'une technologie récente en DFT: le ML. Nous n'avons pas commencé à appliquer le ML pour la densité non locale, cependant, cette technologie montre qu'il pourrait être possible de construire automatiquement des approximations à partir de contraintes physiques. Il est extrêmement difficile pour un théoricien de trouver une fonction mathématique qui satisfait autant de paramètres, donc le ML devient un outil puissant pour accomplir cette tâche. L'ajout d'une nouvelle contrainte au trou X Becke-Roussel à l'aide du ML pourrait être utilisé avec le facteur de corrélation développé par notre groupe de recherche.

Ce projet nous a encouragés à explorer le terme de quatrième ordre du développement en série de puissances de la distance interélectronique du trou X avec le ML. Le terme de quatrième ordre du développement en série de puissances du trou X a été prouvé et obtenu mathématiquement par notre groupe en 2017, mais la construction d'une fonctionnelle est difficile avec cette quantité. Dans cette thèse, ce nouvel ingrédient a donc été combiné avec le ML pour développer une nouvelle approximation qui pourrait permettre d'accélérer les temps de calcul par rapport aux fonctionnelles hybrides qui sont les plus populaires en chimie. Néanmoins, c'est en physique du solide que nos modèles pourraient avoir le plus d'applications, car le calcul de l'échange exacte converge lentement pour des systèmes périodiques. Nos résultats pour les énergies d'atomisation et les barrières d'activation montrent une amélioration considérable des prédictions comparativement aux GGA qui sont très utilisées en physique du solide.

Une contribution importante de ces projets de recherche pour le groupe Ernzerhof, est l'utilisation d'outils libres tels que les logiciels PySCF et PyTorch. Ces outils sont extrêmement puissants et ont été employés pour les projets de cette thèse. PySCF et PyTorch

n'avaient pas été utilisés auparavant par notre groupe, donc les projets de cette thèse ont permis de développer une expertise en ML, ainsi qu'en langage informatique Python. D'ailleurs, comme ces logiciels sont libres, notre groupe de recherche peut aussi distribuer publiquement les codes informatiques de nos modèles: <https://gitlab.com/electronic-structure-udem>. Ce partage de code entre divers groupes de recherche est très bénéfique pour toute la communauté scientifique, nous avons utilisé des codes libres provenant d'autres groupes et nous espérons que nos projets seront aussi utiles pour d'autres chercheurs en DFT.

En résumé, les fonctionnelles qui ont été développées dans cette thèse ne deviendront fort probablement pas les fonctionnelles les plus utilisées en chimie, néanmoins, elles ont contribué à l'avancement des connaissances en DFT. Ce type de recherche fondamental est essentiel pour pouvoir construire de nouvelles fonctionnelles dont les propriétés s'améliorent.

Pour le développement de nouvelles fonctionnelles en DFT, il y a deux philosophies principales. La première est dite non empirique, l'approche préférée de notre groupe de recherche, qui construit de nouvelles fonctionnelles à partir de principes physiques et mathématiques. La deuxième approche est dite empirique, où des paramètres flexibles sont ajustés pour reproduire des bases de données expérimentales ou provenant de calculs de la théorie de la fonction d'onde. Les deux premiers projets de cette thèse sont de la première catégorie, tandis que le troisième est en partie empirique, mais le nombre de molécules pour l'entraînement est inférieur à cinq. Avec le ML, les méthodes empiriques risquent de devenir de plus en plus présentes et les modèles pourront être ajustés pour reproduire des bases de données contenant des milliers de molécules. Quoique cette approche va donner de bonnes prédictions pour diverses propriétés électroniques, les techniques de ML ont généralement des milliers de paramètres ajustables, donc il est très difficile d'analyser le modèle et de comprendre pourquoi de tels résultats sont obtenus. Ainsi, des méthodes non empiriques ou ayant seulement quelques paramètres ajustables sont tout de même essentielles afin de mieux comprendre les principes chimiques et physiques derrière les modèles. Le projet sur les méthodes non locales a montré que des modèles purement non empiriques sont difficiles à combiner avec la densité non locale qui a un comportement fortement oscillatoire, cependant, ce projet a tout de même permis de mieux comprendre la densité non locale, ce qui n'aurait possiblement pas été faisable avec un modèle ML très flexible qui agit comme une boîte noire. Le deuxième et le troisième projet ont utilisé le ML au lieu de construire une équation mathématique simple, puisqu'il est parfois extrêmement complexe de trouver une équation simple pour accomplir une certaine tâche. Le ML permet d'explorer rapidement une quantité comme le terme de quatrième ordre du développement en série de Taylor du trou X.

À court terme, les études réalisées lors de cette thèse ouvrent la possibilité de nouveaux projets comme l'utilisation du ML pour construire un facteur de corrélation. Dans cette thèse nous avons surtout utilisé le ML pour modéliser l'échange, donc la construction d'un facteur de corrélation avec le ML serait une perspective intéressante. Pour le groupe de recherche,

il serait pertinent, par exemple, de construire un facteur de corrélation avec des réseaux de neurones afin de simplifier l'implémentation de la fonctionnelle de manière autocohérente. Les bibliothèques de NN permettent d'obtenir facilement les dérivées des NN, ce qui facilite l'obtention des dérivées du facteur de corrélation qui sont utilisées pour résoudre le champ autocohérent.

De plus, pour un autre projet, le terme de quatrième ordre du développement en série de Taylor du trou X ($T(\mathbf{r})$) est une propriété qui n'a pas beaucoup été étudiée en chimie, donc davantage de travaux sur cette variable pourront permettre de mieux la comprendre afin d'améliorer les fonctionnelles en DFT. L'utilisation de cette variable pour étudier des solides serait particulièrement pertinente, car elle est calculée facilement pour des orbitales obtenues à partir d'onde plane. Pour un projet futur dans le groupe de recherche, il faudra donc calculer $T(\mathbf{r})$ pour des orbitales qui sont formées à l'aide d'onde plane. Ensuite, il faudra trouver des ensembles de données qui sont fiables afin d'entraîner les paramètres des NN. Une fois le NN entraîné, le modèle pourra être appliqué pour calculer les propriétés des solides, comme la largeur de bande, optimiser la structure cristalline du réseau ou étudier des réactions chimiques de surface.

Un autre aspect du ML que nous avons moins mentionné dans cette thèse, mais qui est connecté à nos projets, est l'automatisation des calculs DFT, ainsi que de la robotique en laboratoire (voir cet article de revue par exemple [Kuntz & Wilson. *Pure and Applied Chemistry*, vol. 94, no. 8, 2022, pp. 1019-1054.]) À l'aide de l'intelligence artificielle, il est plus facile d'automatiser certains processus fastidieux comme l'analyse de données massives. Alors, en construisant des fonctionnelles plus efficaces comme nous l'avons fait dans cette thèse, ces fonctionnelles améliorées peuvent être utilisées pour étudier des ensembles de données plus grands. L'étude de grands ensembles de données est pertinente pour trouver de nouveaux médicaments par exemple. Ainsi, une méthodologie qui utilise la DFT pour trouver de nouvelles molécules aux propriétés intéressantes, puis utilise de la robotique en laboratoire afin de les synthétiser et de les tester est extrêmement pertinente pour l'industrie pharmaceutique.

À plus long terme pour la DFT, les nouvelles technologies comme les ordinateurs, le ML, le calcul sur carte graphique et plusieurs autres ont eu un impact considérable sur cette théorie. Il sera intéressant d'assister à l'évolution future de la DFT, qui sera probablement connectée à celle des ordinateurs quantiques, les molécules étant une application particulièrement attrayante pour ces ordinateurs.