Large quantum systems: a mathematical and numerical perspective

Mathieu Lewin

To cite this version:

Mathieu Lewin. Large quantum systems: a mathematical and numerical perspective. Mathematics [math]. Université de Cergy Pontoise, 2009. tel-00394205

HAL Id: tel-00394205
https://theses.hal.science/tel-00394205

Submitted on 10 Jun 2009

Université de Cergy-Pontoise

Systèmes quantiques à grand nombre de particules : une perspective mathématique et numérique

Thèse d'habilitation à diriger des recherches

présentée par

Mathieu LEWIN

JURY

Rapporteurs :	**Volker BACH** Professeur, Johannes Gutenberg-University Mainz (Allemagne)
	Fabrice BETHUEL Professeur, Université Pierre et Marie Curie
	Claude LE BRIS Professeur, École Nationale des Ponts et Chaussées
Examinateurs :	**Maria J. ESTEBAN** Directrice de recherche CNRS, Université Paris-Dauphine
	Vladimir GEORGESCU Directeur de recherche CNRS, Université de Cergy-Pontoise
	Bernard HELFFER Professeur, Université Paris-Sud
	Jan Philip SOLOVEJ Professeur, University of Copenhagen (Danemark)
Coordinateur :	**Éric SÉRÉ** Professeur, Université Paris-Dauphine

9 Juin 2009

Systèmes quantiques à grand nombre de particules : une perspective mathématique et numérique.

Résumé : Ce mémoire est consacré à l'étude mathématique de divers modèles variationnels permettant la description de systèmes quantiques, en particulier infinis. Les outils mathématiques utilisés sont ceux de l'analyse non linéaire, du calcul des variations, des équations aux dérivées partielles, de la théorie spectrale et du calcul scientifique.

Une première partie contient quelques résultats pour des systèmes finis. Nous étudions des approximations de l'équation de Schrödinger pour N électrons dans une molécule ou un atome, puis le modèle de Hartree-Fock-Bogoliubov pour un système de fermions interagissant avec une force de type gravitationnelle.

Dans une seconde partie nous proposons une nouvelle méthode pour démontrer l'existence de la limite thermodynamique pour des systèmes quantiques interagissant avec la force de Coulomb.

Ensuite, nous construisons deux modèles de type Hartree-Fock pour des systèmes infinis. Le premier est un modèle relativiste, déduit de l'électrodynamique quantique, et qui permet de décrire le comportement d'électrons, couplés avec celui du vide de Dirac qui peut se polariser. Le second modèle décrit l'état d'un cristal non relativiste en présence d'un défaut chargé ; il est complété par une nouvelle approche numérique.

La dernière partie du mémoire est consacrée au problème de pollution spectrale, un phénomène observé lorsque l'on cherche à calculer des valeurs propres au milieu du spectre essentiel, par exemple pour des opérateurs de Dirac ou de Schrödinger périodique.

Mots clés : physique mathématique, calcul des variations, équations aux dérivées partielles, théorie spectrale, analyse numérique, mécanique quantique, théorie quantique des champs, chimie quantique, équation de Schrödinger, modèle Hartree-Fock, modèle Hartree-Fock-Bogoliubov, limite thermodynamique, polarisation du vide, renormalisation de charge, défaut dans un cristal, fonctions de Wannier, pollution spectrale.

Large Quantum Systems: a Mathematical and Numerical Perspective.

Abstract: This thesis is devoted to the mathematical study of variational models for large quantum systems. The mathematical methods are that of nonlinear analysis, calculus of variations, partial differential equations, spectral theory, and numerical analysis.

The first part contains some results on finite systems. We study several approximations of the N-body Schrödinger for electrons in an atom or a molecule, and then the so-called Hartree-Fock-Bogoliubov model for a system of fermions interacting via the gravitational force.

In a second part, we propose a new method allowing to prove the existence of the thermodynamic limit of Coulomb quantum systems.

Then, we construct two Hartree-Fock-type models for infinite systems. The first is a relativistic theory deduced from Quantum Electrodynamics, allowing to describe the behavior of electrons, coupled to that of Dirac's vacuum which can become polarized. The second model describes a nonrelativistic quantum crystal in the presence of a charged defect. A new numerical method is also proposed.

The last part of the thesis is devoted to spectral pollution, a phenomenon which is observed when trying to approximate eigenvalues in a gap of the essential spectrum of a self-adjoint operator, for instance for periodic Schrödinger or Dirac operators.

Keywords: mathematical physics, calculus of variation, partial differential equations, spectral theory, numerical analysis, quantum mechanics, quantum field theory, quantum chemistry, Schrödinger equation, Hartree-Fock model, Hartree-Fock-Bogoliubov model, thermodynamic limit, vacuum polarization, charge renormalization, quantum crystal with a defect, Wannier functions, spectral pollution.

CNRS & UNIVERSITÉ DE CERGY-PONTOISE, Laboratoire de Mathématiques UMR CNRS 8088
2, avenue Adolphe Chauvin, 95 302 CERGY-PONTOISE Cedex, FRANCE

Contents

Introduction

Dans ce mémoire, nous présentons divers travaux consacrés à l'étude mathématique de certains systèmes quantiques. Les modèles abordés permettent la description de la matière à l'échelle microscopique, et, surtout, leur comportement lors du passage à l'échelle macroscopique.

La plupart des systèmes étudiés sont composés de fermions interagissant à travers un champ électromagnétique classique (ce dernier n'est pas quantifié et il n'y a donc pas de photon). Certains systèmes sont non relativistes (chapitres 1, 3 et 5) alors que pour d'autres les effets relativistes sont pris en compte (chapitres 2 et 4). Nous étudions aussi des systèmes de fermions interagissant seulement avec la force gravitationnelle. Les modèles étudiés sont issus de la physique atomique et de l'état solide, de la chimie moléculaire et de la théorie quantique des champs. Dans plusieurs cas, ils comprennent un nombre infini de particules.

Les méthodes mathématiques utilisées sont celles de l'analyse non linéaire, du calcul variationnel, des équations aux dérivées partielles, de la théorie spectrale et du calcul scientifique.

Une grande partie du mémoire est consacrée à l'étude de modèles stationnaires par des méthodes variationnelles. Souvent il s'agit de démontrer l'existence d'un minimum (ou de points critiques) pour une fonctionnelle non linéaire modélisant l'énergie (libre) totale du système, et d'établir les propriétés des minimiseurs, pertinentes d'un point de vue physique. Les problèmes sont toujours localement compacts, mais ils sont posés dans tout l'espace physique $\mathbb{R}^3$, avec une possible perte de compacté à l'infini. Dans beaucoup de cas, une originalité notable est que la variable de la fonctionnelle d'énergie est un *opérateur auto-adjoint*, pouvant être de rang infini (contrairement au cadre classique du calcul variationnel où la variable est souvent une fonction à valeurs complexes ou vectorielles). Ceci est toujours le cas lorsque l'on désire modéliser des systèmes quantiques comportant une infinité de particules. Les propriétés des minimiseurs peuvent alors être très délicates et mener à la description de phénomènes physiques subtils (renormalisation de charge au chapitre 4, caractère diélectrique d'un cristal au chapitre 5).

Un problème récurrent dans notre travail est celui de l'étude de certains systèmes dans la limite où le nombre de particules tend vers l'infini. On parle de *limite thermodynamique*. On s'intéresse alors à la convergence de l'énergie (par unité de volume ou par particule) et à l'identification précise de sa limite, ainsi qu'à la convergence des états eux-mêmes. Cette technique est souvent utilisée pour construire des modèles pertinents pour les systèmes infinis.

Nous étudions aussi certaines équations dépendant du temps. Ce sont toujours des systèmes Hamiltoniens (équation de Schrödinger) pour lesquels nous démontrons l'existence de solutions, la stabilité orbitale des solutions stationnaires et, éventuellement, l'explosion en temps fini (si elle a lieu).

Finalement, une dernière composante essentielle de notre travail est la simulation numérique des modèles étudiés, et la construction de méthodes algorithmiques robustes. Pour simuler de tels systèmes, l'effort de calcul à réaliser est souvent déraisonnablement grand. Il est donc très important d'apporter la rigueur nécessaire à l'efficacité des calculs.

Le mémoire est organisé comme suit. Chaque chapitre contient un exposé assez détaillé d'un ensemble de résultats, sur un thème précis. Les chapitres sont quasiment indépendants et peuvent être lus séparément. Le lecteur trouvera ci-après une liste des articles de recherche dont le contenu est développé dans chaque chapitre, avec un très bref résumé.

Chapitre 1. Atomes et molécules non relativistes

Le premier chapitre est consacré à l'étude de l'équation de Schrödinger et de ses approximations non linéaires, pour la description des atomes et des molécules. Un système de N électrons dans une molécule est décrit par un opérateur de Schrödinger ayant la forme

$$H = -\Delta + W(x)$$

et défini sur $L^2((\mathbb{R}^3)^N, \mathbb{C})$ (en fait sur le sous-espace constitué des fonctions antisymétriques). La fonction W comprend l'interaction Coulombienne entre les électrons et les noyaux atomiques (généralement modélisés comme des particules ponctuelles classiques) et l'interaction entre les électrons eux-mêmes. Elle dépend donc des positions spatiales des noyaux atomiques. Le Laplacien modélise lui la somme des énergies cinétiques des N électrons. On s'intéresse aux états liés de la molécule correspondant aux solutions stationnaires de l'équation de Schrödinger, c'est-à-dire aux fonctions propres de H :

$$H\Psi = E\Psi. \tag{1}$$

Si de nombreuses propriétés mathématiques du spectre l'opérateur H sont connues, ceci est souvent d'un bien faible intérêt pratique : la dimension de l'espace $L^2((\mathbb{R}^3)^N, \mathbb{C})$ sur lequel H est défini rend la résolution numérique directe de l'équation (1) tout simplement inaccessible dès que $N \gtrsim 6$ ou 7. Pour cette raison, les physiciens et les chimistes utilisent des modèles approchés, qui sont très souvent non linéaires.

Le premier modèle que nous avons étudié dans [1, 2, 3, 4] est appelé *multi-configurations.* Il consiste à restreindre la forme quadratique $\Psi \mapsto \langle \Psi, H\Psi \rangle$ au sous-ensemble de la sphère de $L^2((\mathbb{R}^3)^N, \mathbb{C})$, constitué des fonctions sous la forme

$$\Psi = \sum_{1 \leq i_1, \cdots i_N \leq K} c_{i_1,\ldots,i_N} \varphi_{i_1} \otimes \cdots \otimes \varphi_{i_N}.$$

Dans cette formule, les $\{\varphi_i\}_{i=1}^K$ forment un système orthonormé de $L^2(\mathbb{R}^3, \mathbb{C})$ et les $\{c_{i_1,\ldots,i_N}\}$ sont des nombres complexes (antisymétriques par rapport aux échanges des i_k). Lorsque $K = N$, on trouve le célèbre modèle *Hartree-Fock* (dont les principales propriétés sont rappelées au chapitre 1 et dont l'équation dépendante du temps a été étudiée dans [8]), alors que lorsque $K \to \infty$ on obtient le modèle initial de Schrödinger posé sur tout l'espace $L^2((\mathbb{R}^3)^N, \mathbb{C})$. Dans [1, 2] nous avons étudié l'existence d'un minimum et de points critiques pour la fonctionnelle obtenue, avec $N \leq K < \infty$. Celle-ci est non-quadratique en les φ_i et les équations d'Euler-Lagrange associées forment un système de K équations aux dérivées partielles elliptiques non linéaires couplées. Nous avons également relié les points critiques construits avec les valeurs propres et les fonctions propres de l'opérateur H. Dans [3], nous proposons une nouvelle méthode numérique pour le premier état excité des électrons (la seconde valeur propre de H), basée sur la formulation variationnelle de [2]. On doit essentiellement calculer un point critique de type col ; des tests sont présentés pour des molécules comprenant deux électrons.

Dans [5], nous avons étudié un autre modèle permettant le calcul approché de la première fonction propre de H, basé sur les *matrices de densité d'ordre deux.* Essentiellement, le problème de minimisation de $\Psi \mapsto \langle \Psi, H\Psi \rangle$ sur $L^2((\mathbb{R}^3)^N, \mathbb{C})$ est remplacé par un problème de minimisation sur un sous-ensemble convexe du plus petit espace $L^2((\mathbb{R}^3)^4, \mathbb{C})$. Comme l'ensemble convexe en question est très difficile à caractériser, il est approché par un convexe plus simple. Plusieurs résultats numériques sont présentés.

Enfin, la dernière partie du chapitre contient un résumé des travaux de [6, 7], dans lesquels est étudié un modèle très simple pour la description des réactions chimiques (celui utilisé par les chimistes pour décrire des réactions adiabatiques). Il consiste à faire varier les positions des noyaux en supposant que les électrons restent toujours dans leur état fondamental (la première fonction propre de H). Une réaction est alors modélisée par un problème variationnel de type "lemme du col" pour la première valeur propre λ_0 de H (qui dépend des positions des noyaux). Si ce problème est posé en dimension finie, montrer l'existence du point col est bien sûr un problème complexe, car il demande de connaître les propriétés précises de $\lambda_0(\cdot)$ 'à l'infini', qui sont elles-mêmes reliées à celles de l'opérateur H.

Publications du chapitre 1

[1] [*]M. Lewin, *The multiconfiguration methods in quantum chemistry : Palais-Smale condition and existence of minimizers*, C. R. Math. Acad. Sci. Paris, 334 (2002), pp. 299–304.

[2] [*]M. Lewin, *Solutions of the multiconfiguration equations in quantum chemistry*, Arch. Ration. Mech. Anal., 171 (2004), pp. 83–114.

[3] [*]É. Cancès, H. Galicher, and M. Lewin, *Computing electronic structures : a new multiconfiguration approach for excited states*, J. Comput. Phys., 212 (2006), pp. 73–98.

[4] M. Lewin, *On the computation of excited states with MCSCF methods*, J. Math. Chem., 44 (2008), pp. 967–980. Conference "Mathematical Methods for Ab Initio Quantum Chemistry", Nice, Nov. 2005.

[5] É. Cancès, M. Lewin, and G. Stoltz, *The electronic ground-state energy problem : a new reduced density matrix approach.*, J. Chem. Phys., 125 (2006), p. 64101.

[6] [*]M. Lewin, *A mountain pass for reacting molecules*, Ann. Henri Poincaré, 5 (2004), pp. 477–521.

[7] M. Lewin, *Solution of a mountain pass problem for the isomerization of a molecule with one free atom*, Ann. Henri Poincaré, 7 (2006), pp. 365–379.

[8] J. Dolbeault, P. Felmer, and M. Lewin, *Stability of the Hartree-Fock model with temperature*, Math. Models Methods Appl. Sci., 19 (2009), pp. 347–367.

Chapitre 2. Le modèle Hartree-Fock-Bogoliubov pour les étoiles à neutrons et les naines blanches

Le chapitre 1 est consacré à l'étude des électrons au sein d'une molécule, c'est à dire de particules quantiques se repoussant (ils sont de même charge) et soumises à un champ extérieur attractif (celui créé par les noyaux). Dans le chapitre 2, nous étudions un modèle non linéaire célèbre pour des particules qui s'attirent et ne sont soumises à aucun champ extérieur. La combinaison entre le caractère attractif de l'interaction et des effets non linéaires permet alors l'existence d'états liés, même si le système est bien sûr globalement invariant par translation.

Le modèle *Hartree-Fock-Bogoliubov* (HFB) étudié dans ce chapitre permet de décrire certaines des propriétés physiques importantes des systèmes quantiques attractifs. Il s'agit d'une fonctionnelle non linéaire dont les variables sont deux opérateurs γ et α, agissant sur $L^2(\mathbb{R}^3, \mathbb{C}^2)$, et reliés par la contrainte :

$$0 \leq \begin{pmatrix} \gamma & \alpha \\ \alpha^* & 1 - \overline{\gamma} \end{pmatrix} \leq 1.$$

Les opérateurs γ et α sont respectivement appelés *matrice densité d'ordre 1* et *matrice d'appariement.* L'opérateur α décrit le comportement des *paires de Cooper*, responsables de la supraconductivité ou de la superfluidité de certains systèmes.

Dans [9], nous avons étudié le modèle HFB pour un système pseudo-relativiste soumis uniquement aux forces gravitationnelles, comme on peut en trouver dans les étoiles à neutrons ou les naines blanches. L'énergie cinétique est maintenant modélisée par l'opérateur pseudo-différentiel

$$T = \sqrt{m^2 - \Delta} - m$$

---

[*] Les publications dont une version (éventuellement préliminaire) était déjà contenue dans ma thèse sont indiquées avec une étoile.

et le potentiel d'attraction est purement gravitationnel. Nous avons démontré l'existence d'un minimum pour la fonctionnelle HFB dès que cette dernière est bornée inférieurement (cela dépend du nombre total de particules) et donné certaines propriétés des minima. Plusieurs difficultés rendent ce problème complexe. Tout d'abord l'énergie est invariante par translation et elle n'est pas semi-continue inférieure pour la topologie faible. Nous avons donc utilisé des techniques de type "concentration-compacité" de P.-L. Lions. Cependant, le fait que l'énergie cinétique est non locale et la difficulté supplémentaire que les variables sont des opérateurs (éventuellement de rang infini) ont constitué des obstacles importants. Enfin, la présence de l'opérateur α constitue une complication notable : le terme d'appariement dans l'énergie est "critique" au sens où il ne peut être contrôlé que par l'énergie cinétique.

Dans une seconde partie, nous énonçons un résultat de [10] : nous considérons l'équation dépendant du temps associée au modèle HFB et nous montrons que si la masse initiale du système est trop grande, le système gravitationnel s'effondre en temps fini, c'est-à-dire la solution temporelle explose en norme. Ceci correspond au comportement que l'on observe pour certaines étoiles trop massives.

Publications du chapitre 2

[9] E. Lenzmann and M. Lewin, *Minimizers for the Hartree-Fock-Bogoliubov theory of neutron stars and white dwarfs.* `arXiv:0809.2560`, 2008.

[10] C. Hainzl, E. Lenzmann, M. Lewin, and B. Schlein. In preparation, 2009.

Chapitre 3. La limite thermodynamique des systèmes quantiques Coulombiens

Dans les deux chapitres précédents, nous avons étudié divers systèmes finis. Le chapitre 3 est le premier de ce mémoire qui est consacré à l'étude de systèmes quantiques infinis.

Un problème naturel est de passer à la limite macroscopique, c'est-à-dire de faire tendre le nombre de particules vers l'infini dans un modèle fini. Cette question est difficile, en particulier lorsque les particules interagissent avec le potentiel de Coulomb, qui est à longue portée (il n'est pas intégrable). C'est principalement grâce à un phénomène d'écrantage (la matière s'organise spontanément de façon à être neutre localement) que la matière usuelle peut exister à l'échelle macroscopique. Quantifier et utiliser cet écrantage pour démontrer la stabilité de systèmes quantiques infinis est complexe.

Si on note $E(N)$ l'énergie fondamentale d'un système comprenant N particules (à définir, bien sûr – voir les détails au chapitre 3), on doit montrer que $E(N) \sim_{N\to\infty} \bar{e}N$ pour une certaine constate $\bar{e}$. En effet imaginons un instant que $E(N) \sim_{N\to\infty} \bar{e}N^a$ pour un certain $a \neq 1$. Si on essaie alors de réunir deux systèmes macroscopiques identiques (deux verres d'eau par exemple), l'énergie à fournir au système sera de l'ordre de $E(2N) - 2E(N) \sim (2^a - 2)\bar{e}N^a$. Comme N est très grand (typiquement $N \sim 10^{23}$), selon le signe de $\bar{e}$ et la position de a par rapport à 1, il faudra soit fournir une énergie colossale pour réunir les deux systèmes (le système serait alors très instable), soit au contraire une énergie gigantesque serait libérée et la réunion de nos deux verres d'eau formerait une sorte de bombe. Comme aucune de ces deux possibilités n'est physiquement raisonnable, on doit donc avoir $E(N) \sim \bar{e}N$.

La démonstration de la convergence de l'énergie par particule $E(N)/N$ quand N tend vers l'infini ainsi que l'identification de la limite constituent un problème crucial pour une meilleure compréhension des systèmes quantiques infinis. Celui-ci a été abondamment étudié depuis la fin des années 60. Dans deux articles [11, 12] résumés dans [13], nous proposons une nouvelle méthode pour prouver l'existence de la limite de l'énergie par particule, basée sur une inégalité électrostatique due à Graf et Schenker et qui sert à quantifier le phénomène d'écrantage. Cette méthode permet de retrouver des résultats précédents de Lieb et Lebowitz, ou de Fefferman de façon unifiée, et également de traiter de nouveaux systèmes, comme cela est expliqué en détails au chapitre 3.

Il est en pratique très difficile d'identifier précisément la limite $\bar{e}$ de l'énergie par unité de volume ou de montrer la convergence des états, sauf pour certains systèmes plus simples. Dans les chapitres 4 et 5, nous étudions deux systèmes avec un modèle de type Hartree-Fock, pour lesquels ce programme peut être mené à bien.

Publications du chapitre 3

[11] C. Hainzl, M. Lewin, and J. P. Solovej, *The thermodynamic limit of quantum Coulomb systems. Part I. General theory*, Advances in Math., 221 (2009), pp. 454–487.

[12] C. Hainzl, M. Lewin, and J. P. Solovej, *The thermodynamic limit of quantum Coulomb systems. Part II. Applications*, Advances in Math., 221 (2009), pp. 488–546.

[13] C. Hainzl, M. Lewin, and J. P. Solovej, *The thermodynamic limit of quantum Coulomb systems : A new approach*, in Mathematical results in Quantum Mechanics : Proceedings of the QMath10 Conference, I. Beltita, G. Nenciu, and R. Purice, eds., World Scientific, 2008.

Chapitre 4. L'approximation de Hartree-Fock en Électrodynamique Quantique sans photon

Le chapitre 4 est probablement le plus long de ce mémoire. Il est consacré à la construction et l'étude détaillée d'un nouveau modèle permettant la description de systèmes relativistes, déduit de l'électrodynamique quantique (une théorie qui a, en grande partie, encore résisté à toute formalisation mathématique non perturbative rigoureuse).

Pour les atomes lourds, il est nécessaire de prendre en compte les effets relativistes, au moins pour les électrons "de coeur" (les plus proches des noyaux). En principe, il suffirait de remplacer l'opérateur $-\Delta$ par l'opérateur de Dirac D^0 (un opérateur différentiel d'ordre un). Or, contrairement à l'opérateur $-\Delta$, le spectre de l'opérateur de Dirac n'est pas borné inférieurement :

$$\sigma(D^0) = (-\infty, -mc^2] \cup [mc^2, \infty).$$

Ceci change radicalement le comportement physique et les propriétés mathématiques du système. Par exemple, en théorie des champs relativiste, le vide n'est plus inerte mais il peut se polariser en présence d'un champ extérieur, cette polarisation est elle même vue par les particules et mène à l'étude d'un système couplé. Si le champ extérieur est très fort, le vide peut même réagir intensément au point de créer spontanément, à partir de l'énergie fournie, une paire électron-positron.

Au chapitre 4, nous étudions l'approximation de Hartree-Fock de l'électrodynamique quantique en négligeant les photons. Le modèle que nous avons obtenu décrit le comportement du vide et ses interactions avec les particules "réelles". Le système physique est toujours composé d'une infinité de particules quantiques : la plupart d'entre elles forment la *mer de Dirac* modélisant le vide alors que seulement un nombre fini d'entre elles sont des particules physiques réelles, comme des électrons ou des positrons.

La première étape est de déterminer l'état du vide libre, c'est-à-dire en l'absence de tout champ extérieur (mais en tenant compte des interactions entre les particules de la mer de Dirac). Pour cela nous avons eu recours dans [18] à une limite thermodynamique comme expliqué plus haut dans notre résumé du chapitre 3 : on commence par poser le modèle sur une boîte de taille L, avec des conditions au bord périodiques, puis on identifie la limite de l'énergie par unité de volume $E^0(L)L^{-3}$. On trouve alors que l'état du système converge vers un état Hartree-Fock comprenant une infinité de particules. Sa matrice densité est un projecteur orthogonal $\mathcal{P}^0_-$ de rang infini, invariant par translation. Même si le vide libre comprend une infinité d'électrons virtuels, sa charge est toujours supposée non observable. Dans les cas les plus simples, le système est simplement représenté par le projecteur spectral négatif $\chi_{(-\infty,0]}(D^0)$. Ici l'opérateur $\mathcal{P}^0_-$, tenant compte des interactions entre les particules, est solution d'une équation non linéaire.

Ensuite, en présence d'un champ extérieur V (par exemple créé par une distribution de charge positive), l'idée est de décrire les modifications du système par rapport au système de référence, le vide libre construit précédemment (voir la figure 1). Nous avons montré dans [18] que lorsque $L \to \infty$, l'énergie fondamentale du système en présence du champ V se comporte comme suit :

$$E^V(L) - E^0(L) \to c.$$

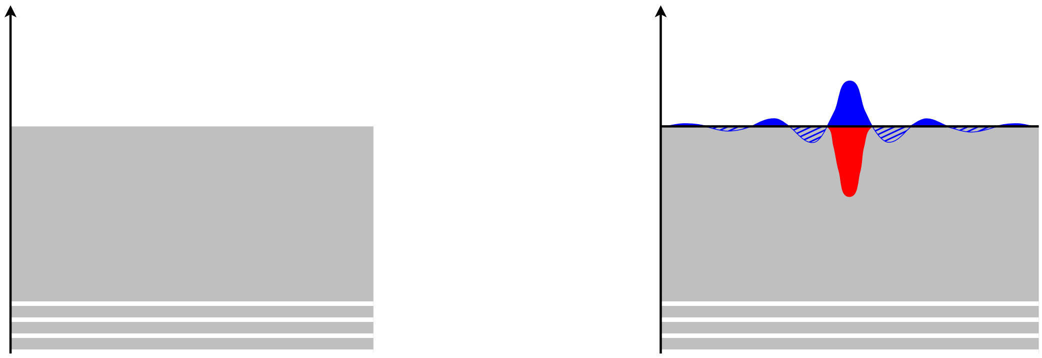
La constante c peut être identifiée comme l'infimum d'une fonctionnelle non linéaire $Q \mapsto \mathcal{E}^V(Q)$ ressemblant à la fonctionnelle Hartree-Fock et dont la variable Q est un opérateur (*a priori* de rang infini) satisfaisant la contrainte $-\mathcal{P}^0_- \leq Q \leq 1 - \mathcal{P}^0_-$. L'opérateur Q décrit les modifications du système en présence de V et on doit avoir $Q = 0$ en l'absence de tout champ ($V = 0$). De même, la différence entre l'état fondamental dans la boîte de taille L et l'état non perturbé converge à la limite vers un minimum Q de la fonctionnelle $\mathcal{E}^V$.

L'existence de minima pour $\mathcal{E}^V$ et leur propriétés ont été étudiées dans [14, 15, 16, 21]. Tout minimiseur est solution d'une équation d'Euler-Lagrange non linéaire sous la forme

$$Q = \chi_{(-\infty,\mu)}(D^0 + V + X_Q) - \mathcal{P}^0_- \tag{2}$$

où X_Q est un opérateur dépendant de Q. L'équation (2) peut s'interpréter comme un système infini d'équations aux dérivées partielles non linéaires couplées. Les principales difficultés pour démontrer l'existence de minima sont : (i) le modèle est posé dans tout l'espace et une perte de compacité à l'infini est possible ; (ii) la variable Q est un opérateur de rang infini soumis à la contrainte $-\mathcal{P}^0_- \leq Q \leq 1 - \mathcal{P}^0_-$; (iii) on s'attend à ce que les minima soient singuliers. En fait, nous avons montré dans [21] que les minima (d'une fonctionnelle simplifiée) ne sont jamais à trace, ce qui est relié au concept de renormalisation de charge. Ceci est expliqué en détails au chapitre 4.

L'équation dépendante du temps et le modèle à température positive sont respectivement étudiés dans [17] et dans [20]. Le chapitre 4 est une version améliorée de [19] et de la dernière partie de [22].



(a) (b)

Fig. 1 – (a) *Dans l'approximation de Hartree-Fock de l'électrodynamique quantique sans photon, le vide est un milieu de charge "infinie mais constante".* (b) *Lorsqu'un potentiel extérieur est appliqué au vide (par exemple si on introduit un noyau chargé, de densité de charge représentée en rouge), ce dernier se polarise. Dans le modèle du chapitre 4, seulement la modification du vide (en bleue) par rapport au vide libre, décrite par l'opérateur Q, est modélisée.*

Publications du chapitre 4

[14] *C. Hainzl, M. Lewin, and É. Séré, *Existence of a stable polarized vacuum in the Bogoliubov-Dirac-Fock approximation*, Commun. Math. Phys., 257 (2005), pp. 515–562.

[15] *C. Hainzl, M. Lewin, and É. Séré, *Self-consistent solution for the polarized vacuum in a no-photon QED model*, J. Phys. A, 38 (2005), pp. 4483–4499.

[16] C. Hainzl, M. Lewin, and É. Séré, *Existence of atoms and molecules in the mean-field approximation of no-photon quantum electrodynamics*, Arch. Rational Mech. Anal., 192 (2009), pp. 453–499.

[17] C. Hainzl, M. Lewin, and C. Sparber, *Existence of global-in-time solutions to a generalized Dirac-Fock type evolution equation*, Lett. Math. Phys., 72 (2005), pp. 99–113.

[18] C. Hainzl, M. Lewin, and J. P. Solovej, *The mean-field approximation in quantum electrodynamics : the no-photon case*, Comm. Pure Appl. Math., 60 (2007), pp. 546–596.

[19] C. Hainzl, M. Lewin, É. Séré, and J. P. Solovej, *A minimization method for relativistic electrons in a mean-field approximation of quantum electrodynamics*, Phys. Rev. A, 76 (2007), p. 052104.

[20] C. Hainzl, M. Lewin, and R. Seiringer, *A nonlinear model for relativistic electrons at positive temperature*, Rev. Math. Phys., 20 (2008), pp. 1283 –1307.

[21] P. Gravejat, M. Lewin, and É. Séré, *Ground state and charge renormalization in a nonlinear model of relativistic atoms*, Commun. Math. Phys., 286 (2009), pp. 179–215.

[22] M. J. Esteban, M. Lewin, and É. Séré, *Variational methods in relativistic quantum mechanics*, Bull. Amer. Math. Soc. (N.S.), 45 (2008), pp. 535–593.

Chapitre 5. L'approximation de Hartree-Fock pour un cristal non relativiste

Nous présentons au chapitre 5 un modèle inspiré de celui présenté au chapitre 4, mais cette fois pour la description des cristaux (non relativistes) avec et sans défauts.

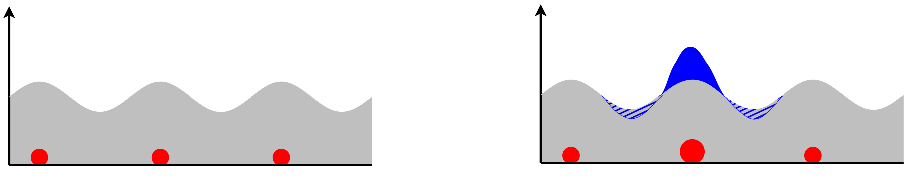
Tout comme le vide de Dirac, un cristal quantique est composé d'une infinité d'électrons formant la *mer de Fermi*. Celle-ci peut aussi se polariser en présence d'un défaut dans le cristal.

Comme précédemment, la première étape est de décrire le cristal parfait périodique. En utilisant une nouvelle fois un procédé de limite thermodynamique, on peut construire un modèle de type Hartree-Fock (réduit) permettant de décrire le système périodique. L'étude des propriétés des minima périodiques a été réalisée dans [25, 24], généralisant des résultats de Catto, Le Bris et Lions.

Ensuite, la seconde étape est l'étude du cristal perturbé, par exemple en présence d'un défaut chargé. Dans [23] nous avons construit et étudié un modèle dans lequel l'état du système est décrit par rapport à l'état "non perturbé" (périodique), voir la figure 2.

Plusieurs des difficultés rencontrées au chapitre 4 pour la description des systèmes relativistes se retrouvent dans l'étude des cristaux non relativistes, avec toutefois des différences notables. En particulier, le Hamiltonien de Schrödinger périodique est cette fois borné inférieurement (contrairement à l'opérateur de Dirac qui a un spectre symétrique par rapport à zéro). Cependant la mer de Fermi possède moins de symétries que la mer de Dirac (qui est invariante par translations) et ceci a des conséquences importantes sur les propriétés des minima, comme prouvé dans [26] et détaillé au chapitre 5.

La dernière partie du chapitre est consacrée à la présentation d'une nouvelle méthode numérique [25] basée sur le modèle variationnel introduit dans [23]. Cette méthode est fondamentalement différente de celle communément utilisée par les chimistes et les physiciens, qui consiste essentiellement à périodiser arbitrairement le système en choisissant une grande boîte bien plus grande que la taille du défaut. Notre méthode, basée plutôt sur une description du modèle en utilisant deux échelles (celle décrivant le réseau périodique et celle décrivant les électrons localisés autour du défaut), semble plus robuste et plus efficace, d'après les premiers tests qui ont été effecutés.



(a) (b)

FIG. 2 – (a) *Les électrons du cristal parfait forment la "mer de Fermi" périodique.* (b) *En présence d'un défaut (un noyau du cristal plus chargé que les autres par exemple), la mer de Fermi se polarise. Dans le modèle de [23] et comme pour le modèle relativiste du chapitre 4, seule la modification par rapport au cristal parfait est modélisée.*

Publications du chapitre 5

[23] É. CANCÈS, A. DELEURENCE, AND M. LEWIN, *A new approach to the modelling of local defects in crystals : the reduced Hartree-Fock case*, Commun. Math. Phys., 281 (2008), pp. 129–177.

[24] M. GHIMENTI AND M. LEWIN, *Properties of periodic Hartree-Fock minimizers*, Calc. Var. Partial Differential Equations, 35 (2009), pp. 39–56.

[25] É. CANCÈS, A. DELEURENCE, AND M. LEWIN, *Non-perturbative embedding of local defects in crystalline materials*, J. Phys. : Condens. Matter, 20 (2008), p. 294213.

[26] É. CANCÈS AND M. LEWIN, *The dielectric permittivity of crystals in the reduced Hartree-Fock approximation.* `arXiv :0903.1944`, 2009.

Chapitre 6. Pollution spectrale et comment l'éviter

Dans le dernier chapitre, nous étudions un phénomène bien connu, sous une perspective nouvelle. Lorsque l'on cherche à calculer le spectre d'un opérateur auto-adjoint A, on a souvent recours à une approximation de type Galerkin : on se fixe un espace de dimension finie et on diagonalise la matrice de la forme quadratique associée à A dans cet espace. On espère ensuite que, lorsque la taille de la base augmente, le spectre converge vers celui de A. Malheureusement, on peut voir assez facilement (un exemple très simple est donné dans l'introduction du chapitre 6) qu'une telle méthode peut fournir, à la limite, des *valeurs propres spécieuses*, ne faisant pas partie du spectre de A. Ceci ne peut se produire que dans des trous spectraux, au milieu du spectre essentiel. On parle alors de *pollution spectrale*. Un tel phénomène est très fréquemment observé lors de calculs sur des opérateurs de Schrödinger périodiques ou des opérateurs de type Dirac, comme ceux rencontrés aux chapitres 4 et 5. Dans ces deux cas, on doit toujours calculer des valeurs propres dans un trou spectral, au milieu du spectre essentiel.

En pratique, on peut soit chercher des critères permettant de savoir si une valeur propre calculée est réelle ou spécieuse, soit au contraire chercher des contraintes sur les espaces vectoriels utilisés pour réaliser le calcul, afin d'éviter complètement le problème de pollution (au moins dans un intervalle donné du spectre). En mécanique quantique relativiste, c'est la seconde solution qui est le plus souvent choisie : les physiciens et chimistes ont introduit divers méthodes dans le but d'éviter le problème de pollution spectrale.

Dans [27], nous étudions le phénomène de pollution spectrale en ajoutant certaines contraintes sur les bases de Galerkin. Nous avons ainsi pu justifier que la méthode numérique proposée dans [25] (pour le calcul des valeurs propres des opérateurs de Schrödinger périodiques avec des bases de Wannier) est exempte de toute pollution. Nous avons également étudié quelques unes des méthodes utilisées en physique et chimie pour les calculs relativistes et déterminé dans quels cas elles sont efficaces et dans quelle situation elles sont théoriquement d'aucun secours.

Publications du chapitre 6

[27] M. Lewin and É. Séré, *Spectral pollution and how to avoid it (with applications to Dirac and periodic Schrödinger operators).* `arXiV:0812.2153`, 2008.

Appendice A. Matrice de densité pour les états de Bogoliubov de charge finie

La dernière partie du mémoire est un appendice dans lequel nous présentons certains résultats abstraits issus de [14, 16] et qui sont constamment utilisés dans les chapitres 2, 4 et 5. Elle fait le lien entre certaines propriétés des matrices densités et le formalisme usuel pour les états de Bogoliubov (ou quasi-libres) en terme d'opérateurs de création et d'annihilation dans un espace de Fock.

Remerciements

J'adresse de vifs remerciements à tous les membres du jury pour leur présence et leur intérêt, en particulier à ceux qui ont accepté d'assumer la charge de rapporteur.

Je tiens aussi à remercier profondément tous mes collaborateurs et amis, avec qui j'ai tant plaisir à travailler et interagir, notamment Éric Séré et Jan Philip Solovej avec qui j'apprends toujours autant.

Merci également à mes collègues et amis de l'université de Cergy-Pontoise, de l'université Paris-Dauphine et de l'École Nationale des Ponts et Chaussées, pour les nombreuses discussions scientifiques ou amicales que nous avons pu avoir.

Introduction

In this thesis, we study, from a mathematical point of view, several quantum systems which are used to describe matter at the microscopic scale, and its behavior when passing to the macroscopic scale.

We will mainly consider fermionic systems interacting through classical electromagnetic forces (as the electromagnetic field is not quantized, there will not be any photon). In some cases we will treat them non-relativistically (Chapter 1, 3 and 5) and in some other cases, we will (maybe partially) take relativistic effects into account (Chapter 2 and 4). We will also study one system only submitted to the gravitational force. All the models that we consider arise from Atomic or Stellar Physics, from Quantum Field Theory, or from Quantum Chemistry. In most cases, our systems will consist of infinitely many quantum particles.

The mathematical methods will be that of nonlinear analysis, calculus of variations, partial differential equations, spectral theory, and numerical analysis.

A large part or our work is focused on the study of stationary states by means of variational methods. The goal will often be to prove the existence of a minimum (or of critical points) for a certain nonlinear energy functional, and to derive some of the properties of minimizers which are the most interesting from a physical point of view. All the problems tackled in this thesis are locally compact (the kinetic energy must stay bounded) but a lack of compactness is always possible at infinity (when the model is posed on the whole space $\mathbb{R}^3$). In many cases, the main variable of our energy will be *a self-adjoint operator*, possibly of infinite rank. This will always be the case when infinitely many particles have to be described. Some properties of the minimizers can then be rather delicate and lead to subtle physical phenomena (for instance charge renormalization in Chapter 4 and dielectric properties of a crystal in Chapter 5).

We will also be often interested in the study of the limit of certain systems when the number of particles goes to infinity (*thermodynamic limit*). In this case we will want to prove the convergence of the energy per particle (or the energy per unit volume) and, possibly, the convergence of states. We will always use this technique to derive suitable models for infinite systems.

Time-dependent equations will also be considered. For the Schrödinger equation corresponding to our energy, we will prove the existence of solutions, the (orbital) stability of stationary states and, possibly, the blow-up in finite-time when it occurs.

Lastly, a large part of our work is devoted to the numerical analysis of our models, and the design of robust algorithms. For the physical systems we are interested in, the computational cost is almost always unreasonably large. It is then very important to use some rigorous insight in order to gain in efficiency.

The thesis is organized as follows. Each chapter contains a rather detailed review of a set of results, on a precise subject. The chapters are almost independent and can be read separately. For each chapter, we will now give a very short summary together with a list of the corresponding publications.

Chapter 1. Nonrelativistic Atoms and Molecules

In the first chapter we study the Schrödinger equation and some of its nonlinear approximations, for the description of atoms and molecules. In a molecule (or an atom), a system of N electrons is described by an N-body Schrödinger operator which has the following form

$$H = -\Delta + W(x).$$

It is defined on $L^2((\mathbb{R}^3)^N, \mathbb{C})$ (indeed on the subspace consisting of antisymmetric wavefunctions). The function W contains both the electrostatic interaction between the electrons and the nuclei, and the self-repulsion between the electrons. The nuclei are usually described as fixed (classical) pointwise particles, hence W depends on their location in $\mathbb{R}^3$. The Laplacian models the sum of the kinetic energy of the electrons. We are interested in the eigenvalues and eigenfunctions of H, solving the time-independent Schrödinger equation

$$H\Psi = E\Psi. \tag{3}$$

Many mathematical properties of H are known, but this is of little practical interest: the dimension of the space $L^2((\mathbb{R}^3)^N, \mathbb{C})$ on which H is defined is so large that it is simply impossible to directly solve numerically Equation (3) when $N \gtrsim 6-7$. For this reason, Physicists and Chemists use many approximate models; they are almost all nonlinear.

The first model that we studied in [1, 2, 3, 4] is called *multiconfiguration.* It consists in restricting the quadratic form associated with H, $\Psi \mapsto \langle \Psi, H\Psi \rangle$, to a smaller set of wavefunctions in the sphere of $L^2((\mathbb{R}^3)^N, \mathbb{C})$, taking the form

$$\Psi = \sum_{1 \leq i_1, \cdots i_N \leq K} c_{i_1,\dots,i_N} \varphi_{i_1} \otimes \cdots \otimes \varphi_{i_N}.$$

In this formula, $\{\varphi_i\}_{i=1}^K$ is an unknown orthonormal system of $L^2(\mathbb{R}^3, \mathbb{C})$ and the $\{c_{i_1,\dots,i_N}\}$ are complex numbers (antisymmetric with respect to exchanges of the indices i_k). When $K = N$, one recovers the celebrated Hartree-Fock model. Its main properties are recalled in Chapter 1; the associated time-dependent equation was studied in [8]. When $K = \infty$, one gets the usual Schrödinger model settled on the whole space $L^2((\mathbb{R}^3)^N, \mathbb{C})$. In [1, 2] we have studied the existence of a minimum and we have constructed specific critical points (for $N \leq K < \infty$) which we have then related to the eigenvalues of H in the limit $K \to \infty$. The energy for $K < \infty$ is non-quadratic in terms of the φ_i's, and the Euler-Lagrange equations form a system of K coupled nonlinear elliptic PDEs. Finally, in [3] we have proposed a new numerical method for calculating an approximation of the first excited state of H, based on the variational formalism of [2]. Essentially, one has to solve a mountain pass problem. Numerical tests for two-electrons systems are provided.

In [5], we have studied another model allowing to compute an approximation of the first eigenvalue of H, based on the formalism of *two-body density matrices.* Loosely speaking, the minimization of the quadratic form $\Psi \mapsto \langle \Psi, H\Psi \rangle$ on the space $L^2((\mathbb{R}^3)^N, \mathbb{C})$ is replaced by a convex minimization problem of an N-dependent linear functional, on a very complicated convex subset of the smaller space $L^2((\mathbb{R}^3)^4, \mathbb{C})$. The latter convex set being very difficult to characterize, it is approximated by a simpler convex set. Several numerical results are provided.

Finally, the last part of Chapter 1 contains a summary of [6, 7], where we have studied a very simple model for chemical reactions (the one which is used by chemists to describe adiabatic reactions). The idea is to vary the position of the nuclei, assuming that the electrons always stay in their ground state (the first eigenfunction of H). A reaction then yields to a mountain pass problem on the first eigenvalue λ_0 of H. Even if the problem is posed in finite dimension (the locations of the nuclei are in $\mathbb{R}^3$), showing the existence of a mountain pass requires to have some information on the 'behavior at infinity' of λ_0, which is of course related to that of H.

---

[*] The papers which were already part of my PhD Thesis (possibly in a preliminary version) are indicated with a star.

References for Chapter 1

[1] *M. Lewin, *The multiconfiguration methods in quantum chemistry: Palais-Smale condition and existence of minimizers*, C. R. Math. Acad. Sci. Paris, 334 (2002), pp. 299–304.

[2] *M. Lewin, *Solutions of the multiconfiguration equations in quantum chemistry*, Arch. Ration. Mech. Anal., 171 (2004), pp. 83–114.

[3] *É. Cancès, H. Galicher, and M. Lewin, *Computing electronic structures: a new multiconfiguration approach for excited states*, J. Comput. Phys., 212 (2006), pp. 73–98.

[4] M. Lewin, *On the computation of excited states with MCSCF methods*, J. Math. Chem., 44 (2008), pp. 967–980. Conference "Mathematical Methods for Ab Initio Quantum Chemistry", Nice, Nov. 2005.

[5] É. Cancès, M. Lewin, and G. Stoltz, *The electronic ground-state energy problem: a new reduced density matrix approach.*, J. Chem. Phys., 125 (2006), p. 64101.

[6] *M. Lewin, *A mountain pass for reacting molecules*, Ann. Henri Poincaré, 5 (2004), pp. 477–521.

[7] M. Lewin, *Solution of a mountain pass problem for the isomerization of a molecule with one free atom*, Ann. Henri Poincaré, 7 (2006), pp. 365–379.

[8] J. Dolbeault, P. Felmer, and M. Lewin, *Stability of the Hartree-Fock model with temperature*, Math. Models Methods Appl. Sci., 19 (2009), pp. 347–367.

Chapter 2. The Hartree-Fock-Bogoliubov Theory of Neutron Stars and White Dwarfs

In Chapter 1 we have studied electrons in a molecule, that is to say fermions with a repulsive force, submitted to an external attractive field created by the nuclei. In Chapter 2, we study a celebrated nonlinear model for systems composed of attractive particles. The attractive feature of the interaction together with some nonlinear effects can then allow the formation of bound states, even in the absence of any external field and although the model is globally invariant by translation.

The *Hartree-Fock-Bogoliubov* (HFB) model which is studied in this chapter allows to describe some important physical properties of attractive systems. It provides us with a nonlinear functional whose variables are two *operators* γ and α acting on $L^2(\mathbb{R}^3, \mathbb{C}^2)$, related by the constraint:

$$0 \leq \begin{pmatrix} \gamma & \alpha \\ \alpha^* & 1-\overline{\gamma} \end{pmatrix} \leq 1.$$

These operators are respectively called the *one-body density matrix* and the *pairing density matrix.* The operator α describes the physical effect of Cooper pairing which is believed to be responsible for the superconductivity or superfluidity of certain physical systems.

In [9], we have studied the HFB model for a system of pseudo-relativistic fermions only submitted to gravitational forces, as may be found for instance in neutron stars or white dwarfs. The kinetic energy is now described by the following *pseudo-relativistic* nonlocal operator

$$T = \sqrt{m^2 - \Delta} - m.$$

We have proved the existence of minimizers for the HFB functional, as soon as it is bounded from below (this depends on the total particle number), and we have derived some important properties of minimizers. We had to face some complications: first the energy is not weakly lower semi-continuous which has led us to use the *concentration-compactness* method of P.-L. Lions. The nonlocality of the kinetic energy operator together with the fact that our variable is a pair of two operators (possibly of infinite rank), linked by a constraint, were important obstacles. Lastly, the operator α cannot be treated by obvious means: the pairing term in the energy is “critical” in the sense that it can only be controlled by the kinetic energy.

In a second part, we state a result of [10] dealing with the associated HFB time-dependent equation and which shows the blow-up in finite-time when the initial datum has a large enough mass (the time-dependent solution then blows up in norm). This corresponds to the observed collapse of certain stars which are too massive.

References for Chapter 2

[9] E. Lenzmann and M. Lewin, *Minimizers for the Hartree-Fock-Bogoliubov theory of neutron stars and white dwarfs.* `arXiv:0809.2560`, 2008.

[10] C. Hainzl, E. Lenzmann, M. Lewin, and B. Schlein. In preparation, 2009.

Chapter 3. The Thermodynamic Limit of Coulomb Quantum Systems

In Chapter 1 and 2, we have studied finite systems. Chapter 3 is the first dealing with infinite quantum systems. A natural problem is to study the *thermodynamic limit*, i.e. the behavior of some (given) system when the number of particles grows. This is a rather involved question, in particular when the particles interact through the Coulomb electrostatic potential which is *long range* (it is not integrable at infinity). It is essentially because of *screening* (matter arranges itself in a way that it is always locally neutral) that usual matter can exist at the macroscopic scale. Quantifying and using this screening in order to show the stability of infinite systems is a challenging problem.

Let us denote by $E(N)$ the ground state energy of a certain quantum system (to be defined, see examples in Chapter 3) with N particles. What we have to prove is that $E(N) \sim_{N\to\infty} \bar{e}N$ for some constant $\bar{e}$. Indeed, let us assume for a while that $E(N) \sim_{N\to\infty} \bar{e}N^a$ for some $a \neq 1$. Then $|E(2N) - 2E(N)|$ becomes very large as $N \gg 1$ (typically $N \sim 10^{23}$). Depending on a and the sign of the constant $\bar{e}$, a very large amount of energy will be either released when two identical systems are put together (like two glasses of water for instance), or necessary to assemble them. As this behavior is not observed in real life, one must have $E(N) \sim_{N\to\infty} \bar{e}N$.

The proof of the convergence of the energy per particle $E(N)/N$ as $N \to \infty$ as well as the identification of the limit is a crucial problem for a better understanding of infinite quantum systems, which has been largely studied since the sixties. In two papers [11, 12] summarized in [13], we have proposed a new method for proving the existence of the thermodynamic limit for systems interacting through Coulomb forces, based on an inequality due to Graf and Schenker, serving as a tool to quantify screening. This method allowed us to recover in a unified way some celebrated results of Lieb and Lebowitz, and of Fefferman, and also to study some other systems which were not known before. Everything is explained in details in Chapter 3.

In practice it is very hard to identify exactly the limit $\bar{e}$ of the energy per particle (or per unit volume), or to prove the convergence of states, except for very simple systems. In Chapter 4 and 5 we study two quantum systems for which this program can be solved completely.

References for Chapter 3

[11] C. Hainzl, M. Lewin, and J. P. Solovej, *The thermodynamic limit of quantum Coulomb systems. Part I. General theory*, Advances in Math., 221 (2009), pp. 454–487.

[12] C. Hainzl, M. Lewin, and J. P. Solovej, *The thermodynamic limit of quantum Coulomb systems. Part II. Applications*, Advances in Math., 221 (2009), pp. 488–546.

[13] C. Hainzl, M. Lewin, and J. P. Solovej, *The thermodynamic limit of quantum Coulomb systems: A new approach*, in Mathematical results in Quantum Mechanics: Proceedings of the QMath10 Conference, I. Beltita, G. Nenciu, and R. Purice, eds., World Scientific, 2008.

Chapter 4. No-Photon Quantum Electrodynamics in the Hartree-Fock Approximation

In Chapter 4 (probably the longest of this thesis), we derive and study a model allowing to describe relativistic electrons. Our model is deduced from Quantum Electrodynamics (QED), a theory which, despite its incredible predictive power, is not yet really understood from a mathematical point of view.

For heavy atoms, it is necessary to take relativistic effects into account, at least for the core electrons in an atom (those close to the nucleus). In principle, it suffices to replace the nonrelativistic kinetic energy operator $-\Delta$ by the Dirac operator D^0 (a differential operator of order one). Contrarily to the Laplacian, the spectrum of D^0 is not bounded from below,

$$\sigma(D^0) = (-\infty, -mc^2] \cup [mc^2, \infty),$$

a property which has many physical consequences. For instance, in QED the vacuum is never inert but it can react to an external field and become polarized. This polarization is itself seen by the particles, leading to the study of a coupled system. If the field is large enough, the vacuum can even react so strongly that an electron-positron pair can be spontaneously created.

In Chapter 4, we study the Hartree-Fock approximation of QED, neglecting photons. Our model is able to describe the behavior of the vacuum coupled to that of the real particles. The system always contains infinitely many particles: most of them belong to the vacuum (the so-called *Dirac sea*), and only finitely many are 'real' particles like electrons or positrons.

The first step is to construct the free vacuum, i.e. the Dirac sea in the absence of any external field, taking into account all interactions between particles. For this, we used in [18] a thermodynamic limit procedure, as was explained in the previous paragraph devoted to Chapter 3. One starts by defining the model in a box of size L, with periodic boundary conditions, and one studies its limit as $L \to \infty$. It was found that the free vacuum is a Hartree-Fock state containing infinitely many particles whose one-body density matrix is an orthogonal projector $\mathcal{P}^0_-$, of infinite rank, invariant by translations. Even if the vacuum contains infinitely charged particles, its charge is itself supposed to be physically unobservable. In the simplest case, the vacuum is described by the negative spectral orthogonal projector $\chi_{(-\infty,0)}(D^0)$. In our case, it is a more complicated projector solution of a nonlinear equation, due to fact that we do not neglect the interactions between the particles.

In the presence of an external field V (for instance induced by the positive distribution of charge of a nucleus), the main idea is then to describe the change of the vacuum measured respectively to the reference system, the free vacuum (see Figure 1). We proved in [18] that when $L \to \infty$, the ground state energy $E^V(L)$ of the system in the box with the external field V behaves as follows

$$E^V(L) - E^0(L) \to c.$$

The constant c can be identified as the minimum of some nonlinear functional $Q \mapsto \mathcal{E}^V(Q)$ having the same form as the Hartree-Fock energy. However Q is now a (possibly infinite-rank) operator defined on the whole space and satisfying the constraint $-\mathcal{P}^0_- \leq Q \leq 1 - \mathcal{P}^0_-$. The model is such that when $V = 0$, one has $Q = 0$. Similarly, the difference between the ground states with and without V converges as $L \to \infty$ to a minimizer Q of the functional $\mathcal{E}^V$.

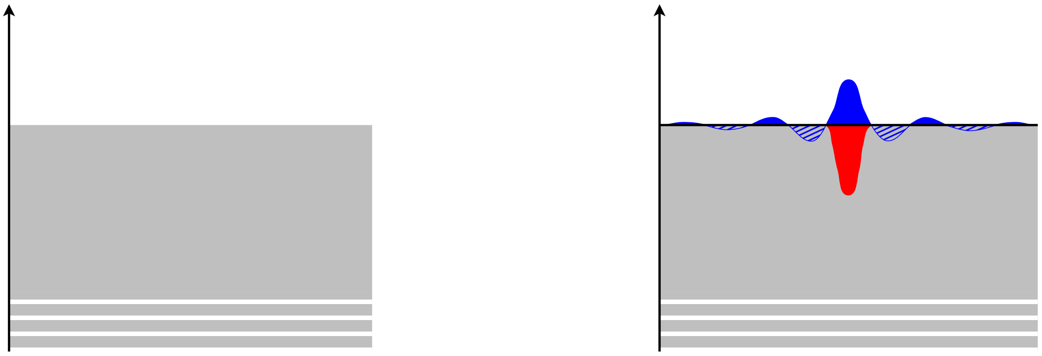
The existence of a minimum for $\mathcal{E}^V$ and the properties of minimizers were studied in [14, 15, 16, 21]. Any minimizer is solution to a nonlinear Euler-Lagrange equation taking the form

$$Q = \chi_{(-\infty,\mu)}(D^0 + V + X_Q) - \mathcal{P}^0_-$$

where X_Q is an operator depending on Q. This equation can be viewed as an infinite system of coupled nonlinear PDEs.

The main difficulties in proving the existence of minimizers are as follows: (i) the model is posed on the whole space and a lack of compactness at infinity is possible; (ii) the variable Q is an operator of infinite rank submitted to the constraint $-\mathcal{P}^0_- \leq Q \leq 1 - \mathcal{P}^0_-$; (iii) minimizers are believed to be rather singular. Indeed we have shown for a simplified model in [21] that minimizers are *never trace-class*. This is itself related to charge renormalization, as we will explain in details in Chapter 4.

The time-dependent equation and the model at positive temperature were respectively studied in [17] and [20]. Chapter 4 is an enhanced version of [19] and of the last part of [22].



(a) (b)

Figure 1: (a) *In the Hartree-Fock approximation of no-photon QED, the free vacuum is a translation-invariant medium with infinite charge.* (b) *In the presence of an external potential V, for instance induced by a positive charge distribution as displayed in red, the vacuum becomes polarized. In our model of Chapter 4, only the modification with respect to the free vacuum (in blue), modelled by the operator Q, is described.*

References for Chapter 4

[14] *C. Hainzl, M. Lewin, and É. Séré, *Existence of a stable polarized vacuum in the Bogoliubov-Dirac-Fock approximation*, Commun. Math. Phys., 257 (2005), pp. 515–562.

[15] *C. Hainzl, M. Lewin, and É. Séré, *Self-consistent solution for the polarized vacuum in a no-photon QED model*, J. Phys. A, 38 (2005), pp. 4483–4499.

[16] C. Hainzl, M. Lewin, and É. Séré, *Existence of atoms and molecules in the mean-field approximation of no-photon quantum electrodynamics*, Arch. Rational Mech. Anal., 192 (2009), pp. 453–499.

[17] C. Hainzl, M. Lewin, and C. Sparber, *Existence of global-in-time solutions to a generalized Dirac-Fock type evolution equation*, Lett. Math. Phys., 72 (2005), pp. 99–113.

[18] C. Hainzl, M. Lewin, and J. P. Solovej, *The mean-field approximation in quantum electrodynamics: the no-photon case*, Comm. Pure Appl. Math., 60 (2007), pp. 546–596.

[19] C. Hainzl, M. Lewin, É. Séré, and J. P. Solovej, *A minimization method for relativistic electrons in a mean-field approximation of quantum electrodynamics*, Phys. Rev. A, 76 (2007), p. 052104.

[20] C. Hainzl, M. Lewin, and R. Seiringer, *A nonlinear model for relativistic electrons at positive temperature*, Rev. Math. Phys., 20 (2008), pp. 1283 –1307.

[21] P. Gravejat, M. Lewin, and É. Séré, *Ground state and charge renormalization in a nonlinear model of relativistic atoms*, Commun. Math. Phys., 286 (2009), pp. 179–215.

[22] M. J. Esteban, M. Lewin, and É. Séré, *Variational methods in relativistic quantum mechanics*, Bull. Amer. Math. Soc. (N.S.), 45 (2008), pp. 535–593.

Chapter 5. Nonrelativistic Crystal in the Hartree-Fock Approximation

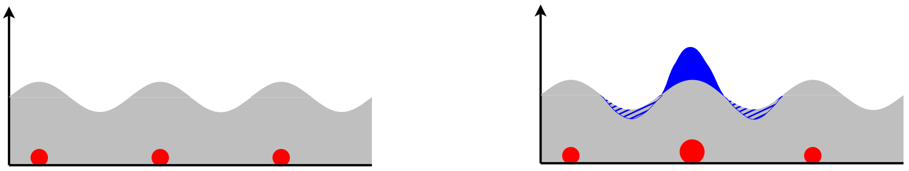
In Chapter 5, we present a model for nonrelativistic crystals with defects, which was inspired of the relativistic model of Chapter 4. Like for Dirac's vacuum, a (nonrelativistic) quantum crystal is composed of infinitely many electrons forming the *Fermi sea*. The latter can also become polarized in the presence of a defect in the crystal.

As before, the first step is to describe the perfect (periodic) crystal. Using again a thermodynamic limit procedure, it is possible to derive a (reduced) Hartree-Fock model for the infinite periodic system. The study of the minimizers was tackled in [25, 24], complementing results of Catto, Le Bris and Lions.

The second step is the study of the perturbed crystal, for instance in the presence of a (localized) charged defect. In [23] we have constructed a model in which the state is described by its modification with respect to the unperturbed, periodic, Fermi sea (see Figure 2), similarly to the QED case.

Several of the difficulties encountered in Chapter 4 for the description of relativistic systems arise similarly in the study of crystals. There are some important differences. In particular, the periodic Schrödinger Hamiltonian is bounded below, contrarily to the Dirac operator. However the crystalline periodic Fermi sea has much less symmetries than the free Dirac sea which is invariant by translations. This has important consequences on the properties of minimizers, as was proved in [26] and will be explained in Chapter 5.

In the last part of Chapter 5, we detail a new numerical method of [25] based on the variational model of [23]. It is very different from the one which is used by Chemists and Physicists which essentially consists in arbitrarily periodizing the system in a box much larger than the size of the defect. Our method, on the contrary, is based on a two-scale description of the crystal (the periodic sea and the localized modifications due to the defect). Our first tests seem to indicate that it is more efficient and robust.



(a) (b)

Figure 2: (a) *The electrons of the perfect crystal form the periodic* Fermi sea. (b) *In the presence of a defect (for instance induced by a nuclei with a higher charge), the Fermi sea becomes polarized. In our model of Chapter 5 and like in Chapter 4, only the modification with respect to the reference state (the periodic Fermi sea) is described.*

References for Chapter 5

[23] É. Cancès, A. Deleurence, and M. Lewin, *A new approach to the modelling of local defects in crystals: the reduced Hartree-Fock case*, Commun. Math. Phys., 281 (2008), pp. 129–177.

[24] M. Ghimenti and M. Lewin, *Properties of periodic Hartree-Fock minimizers*, Calc. Var. Partial Differential Equations, 35 (2009), pp. 39–56.

[25] É. Cancès, A. Deleurence, and M. Lewin, *Non-perturbative embedding of local defects in crystalline materials*, J. Phys.: Condens. Matter, 20 (2008), p. 294213.

[26] É. Cancès and M. Lewin, *The dielectric permittivity of crystals in the reduced Hartree-Fock approximation.* `arXiv:0903.1944`, 2009.

Chapter 6. Spectral Pollution and How To Avoid It

The last chapter is devoted to the study of an important, well-known effect, in a rather new perspective. When trying to compute the spectrum of a self-adjoint operator A, one usually uses a Galerkin-type approximation: a finite-dimensional space is fixed and the matrix of the quadratic form of A in this space is diagonalized. It is hoped that the so-obtained spectrum will converge to the spectrum of A when the size of the basis is enlarged. Unfortunately, it can be seen on very simple examples (see, e.g., the introduction of Chapter 6) that such a method can provide *spurious eigenvalues* which have nothing to do with the true spectrum of A. This phenomenon, which can only happen in gaps of the essential spectrum, is called *spectral pollution.* It is very often observed when trying to calculate the spectrum of periodic Schrödinger operators or of Dirac operators, like those we will meet in Chapters 4 and 5 (in both cases, we are indeed interested in computing eigenvalues in gaps of the essential spectrum).

In practice, one can either try to find a criterion allowing to decide if a computed eigenvalue is real or spurious, or one can look for constraints on the basis which will avoid the phenomenon of pollution, at least in a fixed interval of the spectrum. In relativistic calculations, it is the second option which has been chosen by Chemists and Physicists. They have introduced several methods in order to avoid spectral pollution (like the so-called *kinetic balance*).

In [27], we have studied from a rather abstract point of view the phenomenon of spectral pollution, when some very natural constraints are added on the Galerkin basis. This allowed us in particular to show that the method of [25], reviewed in Chapter 5 (using Wannier functions for approximating the spectrum of periodic Schrödinger operators) does not produce any spurious eigenvalue. We also have studied the methods used in Quantum Chemistry and Physics for relativistic calculations. We have rigorously decided when they are efficient or when they are theoretically of no help.

References for Chapter 6

[27] M. Lewin and É. Séré, *Spectral pollution and how to avoid it (with applications to Dirac and periodic Schrödinger operators).* `arXiV:0812.2153`, 2008.

Appendix A. Density Matrices of Bogoliubov States Having Finite Charge

The last part of the thesis is an appendix in which we present some abstract results of [14, 16],which we constantly use in Chapters 2, 4 and 5. It clarifies the links between some properties of density matrices and the formalism of Bogoliubov (quasi-free) states in Fock space.

Acknowledgments

I am grateful to the members of the jury for their presence and their interest, in particular to those who accepted to write a report on the present manuscript.

I also would like to thank all my collaborators and friends with whom working and interacting was so pleasant. In particular, I deeply thank Éric Séré and Jan Philip Solovej from whom I always learn a lot.

My thanks also go to my colleagues and friends from the University of Cergy-Pontoise, the University of Paris-Dauphine and the École Nationale des Ponts et Chaussées, for numerous scientific or friendly discussions.

1

Nonrelativistic atoms and molecules

In this first chapter, we review some results on *finite systems* like atoms and molecules. In Section 1.1, we recall the properties of the N-body Schrödinger model and of the Hartree-Fock approximation. Section 1.2 (multiconfiguration methods) and 1.3 (Reduced 2-body density matrices) are devoted to two models for approximating the eigenvalues of the linear Schrödinger operator, for fixed positions of the nuclei. They contain numerical results. Lastly, in Section 1.4 we consider a mountain pass problem modeling chemical reactions, and in which the position of the nuclei is not fixed anymore.

1.1 N-body Coulomb quantum systems

1.1.1 The N-body Coulomb Hamiltonian

We consider a molecule containing N non relativistic electrons and M nuclei of charges $Z_1, ..., Z_M$. The nuclei are supposed to be correctly described by a classical model (Born-Oppenheimer approximation) and are thus represented as pointwise charges at $R_1, ..., R_M \in \mathbb{R}^3$. In what follows, we let

$$R = (R_1, ..., R_M) \in (\mathbb{R}^3)^M \setminus (\cup_{i\neq j}\{R_i = R_j\})$$

and

$$Z = (Z_1, ..., Z_M) \in (\mathbb{R}^+)^M, \quad |Z| = Z_1 + \cdots + Z_M.$$

The system is described by the purely Coulombic N-body Hamiltonian

$$H^N(R,Z) = \sum_{i=1}^N \left(-\frac{1}{2}\Delta_{x_i} + V_{R,Z}(x_i)\right) + \sum_{1\leq i<j\leq N} \frac{1}{|x_i - x_j|} + \sum_{1\leq i<j\leq M} \frac{Z_i Z_j}{|R_i - R_j|},$$

where $V_{R,Z}$ is the electrostatic potential created by the nuclei:

$$V_{R,Z}(u) = -\sum_{j=1}^M \frac{Z_j}{|u - R_j|}.$$

As electrons are fermions they must satisfy Pauli's principle which is mathematically expressed by restricting the Hamiltonian $H^N(R,Z)$ to the subspace $\bigwedge_1^N L^2(\mathbb{R}^3,\mathbb{C})$ of $\bigotimes_1^N L^2(\mathbb{R}^3,\mathbb{C})$ consisting of *antisymmetric* wavefunctions:

$$\forall \sigma \in S_N,\ \ \Psi(x_1, ..., x_N) = \epsilon(\sigma)\Psi(x_{\sigma(1)}, ..., x_{\sigma(N)}).$$

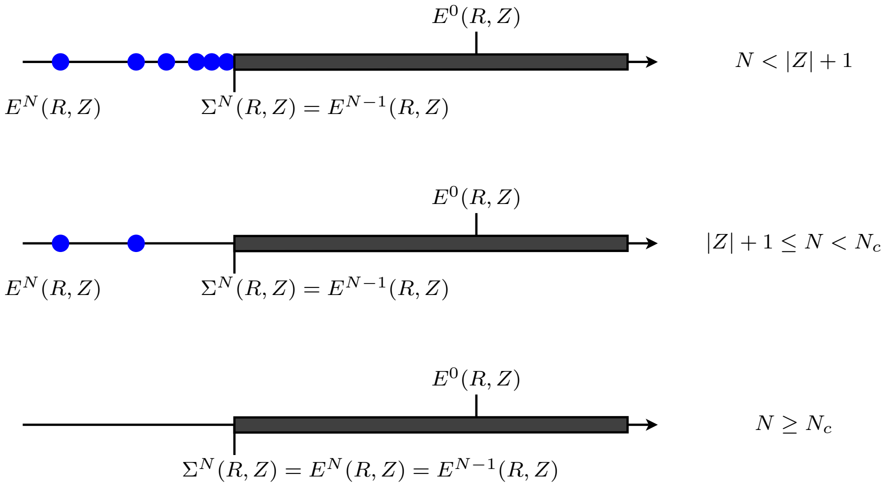


Figure 1.1: Discrete spectrum of $H^N(R,Z)$ depending on the number of electrons.

The quantum energy of the system in a state $\Psi\in\bigwedge_1^N H^1(\mathbb{R}^3,\mathbb{C})$ is the associated quadratic form

$$\mathcal{E}^N(R,\Psi)=\left\langle\Psi,H^N(R,Z)\Psi\right\rangle.$$

For simplicity, we have neglected the spin variable. We work within the so-called *atomic units* for which the mass and the charge of the electrons are both set to one.

For any fixed positions R of the nuclei, it is known [38] that $H^N(R,Z)$ is self-adjoint on the Sobolev space $\bigwedge_1^N H^2(\mathbb{R}^3,\mathbb{C})$, with form domain $\bigwedge_1^N H^1(\mathbb{R}^3,\mathbb{C})$. Also $H^N(R,Z)$ is bounded from below, hence the ground state energy

$$E^N(R,Z):=\inf\sigma\left(H^N(R,Z)\right) \tag{1.1}$$

is a finite quantity. The essential spectrum of $H^N(R,Z)$ takes the form $[\Sigma^N(R,Z),\infty)$. Indeed the *HVZ Theorem* [35, 82, 89, 69] states that

$$\Sigma^N(R,Z)=E^{N-1}(R,Z) \tag{1.2}$$

with the convention that

$$E^0(R,Z)=\sum_{1\leq i<j\leq M}\frac{Z_iZ_j}{|R_i-R_j|}.$$

Below the essential spectrum, there can be some eigenvalues, depending on the value of N and $|Z|$ (see Fig. 1.1). The lowest eigenvalue is called the *ground state energy* and any associated eigenfunction is a *ground state*. Higher eigenvalues describe *excited states*. When there are too many electrons compared to the number of nuclei, intuitively there should not be any bound state. It is known [90, 91] that when $N<|Z|+1$ (neutral or positively charged molecules), there are *infinitely many* eigenvalues converging to the bottom of the essential spectrum $\Sigma^N(R,Z)$. When $N\geq|Z|+1$, the Hamiltonian $H^N(R,Z)$ has *finitely many* eigenvalues below the essential spectrum [87, 83, 74] and there exists a critical N_c such that for $N\geq N_c$, $H^N(R,Z)$ has *no eigenvalue* below its essential spectrum [71, 74, 75]. Lieb proved in [51] that $N_c\leq 2|Z|+M$.

In this chapter we will always denote by $\lambda_d^N(R,Z)$, $d\geq 0$, the ordered eigenvalues of $H^N(R,Z)$, with the convention that $\lambda_d^N(R,Z)=\Sigma^N(R,Z)$ if there are less than d eigenvalues. Note that $\lambda_0^N(R,Z)=E^N(R,Z)$.

In Quantum Chemistry, one is interested in computing an approximation of the different eigenvalues and eigenfunctions, in particular when the positions of the nuclei R varies (chemical reactions). The Schrödinger equation is a model of extremely high accuracy, except for heavy atoms for which core electrons are relativistic. For systems involving a few (say today six or seven) electrons, a direct Galerkin discretization is possible; such a technique is referred to as *Full CI* in Computational Chemistry. For larger systems, this direct approach is out of reach, due to the excessive dimension of the space $\mathbb{R}^{3N}$ on which the wavefunctions are defined, and some approximation must be used. To date, the most commonly used approximations are the Hartree-Fock model (described below in Section 1.1.2) on the one hand, and the Kohn-Sham model (see e.g. [39, 24]) on the other hand. Both of them have been designed for the calculation of ground states and are not really adapted to the calculation of excited states. On the contrary, the MCSCF approximation which will be described in Section 1.2 can be applied to both ground and excited state calculations.

In the next sections, we present several models aiming at calculating approximations of the eigenfunctions and eigenvalues of $H^N(R,Z)$. Only in Section 1.4 we will vary the positions of the nuclei.

1.1.2 Hartree-Fock theory

In this section, we briefly recall the properties of the Hartree-Fock approximation, which is at the basis of many of the works presented here.

In the Hartree-Fock (HF) approximation, one computes an approximation of the first eigenvalue $E^N(R,Z)$ of $H^N(R,Z)$ by restricting the quadratic form $\Psi \mapsto \langle \Psi, H^N(R,Z)\Psi\rangle$ to the class of the functions Ψ which are a simple (so-called *Slater*) determinant:

$$\Psi = \varphi_1 \wedge \cdots \wedge \varphi_N \tag{1.3}$$

where $(\varphi_1, ..., \varphi_N)$ is an orthonormal system of $L^2(\mathbb{R}^3, \mathbb{C})$, $\int_{\mathbb{R}^3} \overline{\varphi_i}\varphi_j = \delta_{ij}$. Equation (1.3) means more precisely

$$\Psi(x_1, ..., x_N) = \frac{1}{\sqrt{N!}} \det(\varphi_i(x_j)).$$

Since the set of all the Ψ's having the form (1.3) is not a vector subspace of $\bigwedge_{i=1}^N L^2(\mathbb{R}^3, \mathbb{C})$, one then obtains an energy functional which is nonlinear in terms of $\varphi_1, ..., \varphi_N$.

For any wavefunction $\Psi \in \bigwedge_1^N L^2(\mathbb{R}^3, \mathbb{C})$, it is convenient to define the *one-body density matrix* γ_Ψ associated with Ψ. This is a self-adjoint trace-class operator acting on $L^2(\mathbb{R}^3, \mathbb{C})$ such that $0 \leq \gamma_\Psi \leq 1$ and $\mathrm{Tr}(\gamma_\Psi) = N$, whose kernel is defined by

$$\gamma_\Psi(x,y) = N \int_{\mathbb{R}^3} \cdots \int_{\mathbb{R}^3} \overline{\Psi(x, x_2, ..., x_N)} \Psi(y, x_2, ..., x_N)\, dx_2 \cdots dx_N. \tag{1.4}$$

The associated *density of charge* is defined by

$$\rho_\Psi(x) := \gamma(x,x) = N \int_{\mathbb{R}^3} \cdots \int_{\mathbb{R}^3} |\Psi(x, x_2, ..., x_N)|^2\, dx_2 \cdots dx_N. \tag{1.5}$$

For a Hartree-Fock state, i.e. when Ψ takes the special form (1.3), one finds that γ_Ψ is precisely the orthogonal projector on the N-dimensional space spanned by $(\varphi_1, ..., \varphi_N)$:

$$\gamma_\Psi = \sum_{i=1}^N |\varphi_i\rangle\langle\varphi_i|, \qquad \rho_\Psi(x) = \sum_{i=1}^N |\varphi_i(x)|^2. \tag{1.6}$$

The energy of a HF state can then be computed and one gets

$$\left\langle \Psi, H^N(R,Z)\Psi \right\rangle = \mathrm{Tr}\left(\left(-\frac{\Delta}{2} + V_{R,Z}\right)\gamma_\Psi\right) + \frac{1}{2}D(\rho_\Psi, \rho_\Psi) - \frac{1}{2}\iint_{\mathbb{R}^3\times\mathbb{R}^3} \frac{|\gamma_\Psi(x,y)|^2}{|x-y|} dx\, dy + E^0(R,Z) \tag{1.7}$$

where we have used the notation

$$D(f,f) := \iint_{\mathbb{R}^3\times\mathbb{R}^3} \frac{f(x)f(y)}{|x-y|}dx\,dy.$$

As can be noticed, the energy only depends on the rank-N projector γ_Ψ (hence on the N-dimensional subspace spanned by $\varphi_1,...,\varphi_N$) and not on the φ_i's themselves. Assuming the nuclei are fixed, we can discard the self-interaction between them and introduce the Hartree-Fock energy functional

$$\mathcal{E}_{\rm HF}(\gamma) := \operatorname{Tr}\left(\left(-\frac{\Delta}{2}+V_{R,Z}\right)\gamma\right) + \frac{1}{2}D(\rho_\gamma,\rho_\gamma) - \frac{1}{2}\iint_{\mathbb{R}^3\times\mathbb{R}^3} \frac{|\gamma(x,y)|^2}{|x-y|}dx\,dy \tag{1.8}$$

where γ is given by (1.6), i.e. it is a projector of rank N. The last two terms of (1.8) are respectively called the *direct* and *exchange terms.*

The Hartree-Fock ground state energy reads

$$E^N_{\rm HF}(R,Z) := \inf_{\substack{\gamma^2=\gamma\\ \operatorname{Tr}(\gamma)=N}} \mathcal{E}_{\rm HF}(\gamma) \tag{1.9}$$

and one of course has $E^N_{\rm HF}(R,Z)+E^0(R,Z) \geq E^N(R,Z)$ (the inequality is even strict [43]). Existence of a minimizer for (1.9) when $N < |Z|+1$ was proved first by Lieb and Simon in [52], and then with a different method by Lions in [56], who also proved the existence of infinitely many critical points. Like in the Schrödinger theory, there is a critical $N_c^{\rm HF}$ above which the Hartree-Fock energy functional has no minimizer anymore. In the atomic case $M=1$, it has been shown by Solovej in [78] that $N_c \leq |Z|+C$ for some constant C. The *reduced Hartree-Fock* model in which the (only nonconvex) exchange term is neglected was studied before in [77]. The associated energy reads:

$$\mathcal{E}_{\rm rHF}(\gamma) := \operatorname{Tr}\left(\left(-\frac{\Delta}{2}+V_{R,Z}\right)\gamma\right) + \frac{1}{2}D(\rho_\gamma,\rho_\gamma) \tag{1.10}$$

The Euler-Lagrange equations for a critical point of $\mathcal{E}_{\rm HF}$ form a system of N coupled nonlinear Partial Differential Equations:

$$H_{\gamma_\Psi}\varphi_k = \lambda_k\varphi_k, \quad k=1,...,N \tag{1.11}$$

where H_{γ_Ψ} is the so-called *mean-field operator* seen by each of the N electrons

$$H_{\gamma_\Psi} = -\frac{\Delta}{2} + V_{R,Z} + \rho_\Psi * \frac{1}{|\cdot|} - \frac{\gamma_\Psi(x,y)}{|x-y|}. \tag{1.12}$$

Note that we have defined the last operator of (1.12) through its kernel. The λ_k's appearing in (1.11) are Euler-Lagrange multipliers arising from the constraint $\langle\varphi_i,\varphi_j\rangle = \delta_{ij}$.

An important result which was shown by Lieb in [50] is that one can actually relax the constraint $\gamma^2=\gamma$ in (1.9) and still obtain the same minimum:

$$E^N_{\rm HF}(R,Z) := \inf_{\substack{0\leq\gamma\leq 1\\ \operatorname{Tr}(\gamma)=N}} \mathcal{E}_{\rm HF}(\gamma). \tag{1.13}$$

This can be interpreted in terms of generalized Hartree-Fock states [5]; more details will be given later in Chapter 2. We note that the equality (1.13) has been proved to be important for deriving well-behaved algorithmic methods [11, 12, 8, 40, 9]. Also it was proved in [4, 5] that the $(N+1)st$ eigenvalue of the operator H_{γ_Ψ} appearing in (1.11) is always above the last filled level: $\lambda_{N+1} > \lambda_N$. This means that the coupled system of PDEs (1.11) may be written in the simple form

$$\boxed{\gamma_\Psi = \chi_{(-\infty,\lambda_N]}(H_{\gamma_\Psi})} \tag{1.14}$$

where χ_I is the characteristic function of the interval I and $\chi_I(H)$ is the associated spectral projector of H.

In Kohn-Sham models [39, 24, 42, 2] an energy of the same form as (1.8) is considered

$$\mathcal{E}_{\rm KS}(\gamma) := \operatorname{Tr}\left(\left(-\frac{\Delta}{2}+V_{R,Z}\right)\gamma\right)+\frac{1}{2}D(\rho_\gamma,\rho_\gamma)+F^{\rm xc}(\rho_\gamma) \tag{1.15}$$

where $F^{\rm xc}$ is a function of the density ρ_γ only and which aims at approximating both the exchange term and the *correlation energy* which is just the difference between the Hartree-Fock and the (true) Schrödinger energy $E^N(R,Z)$.

Let us mention that in this chapter we will mainly be interested in approximating the stationary Schrödinger equation, i.e. in computing eigenstates. There is also a time-dependent Hartree-Fock equation which is a Hamiltonian system associated with the above energy $\mathcal{E}_{\rm HF}$:

$$i\frac{\partial}{\partial_t}\gamma = [H_\gamma\,,\,\gamma]. \tag{1.16}$$

Any minimizer, solution of (1.14), can be shown to yield an *orbitaly stable* stationary solution of (1.16). In a work with J. Dolbeault and P. Felmer [22], we have given a method to construct a large class of orbitaly stable stationary generalized HF states by minimizing free energy-like functionals of the form

$$\gamma \mapsto \mathcal{E}_{\rm HF}(\gamma)+\operatorname{Tr}(\beta(\gamma)),$$

where β is an adequate convex function on $[0,1]$. The usual free energy at temperature T would correspond to $\beta(\nu)=T(\nu\log\nu+(1-\nu)\log(1-\nu))$ but it does not yield a bounded below functional on the whole space. Considering other functions (for instance $\beta(\nu)=\nu^m$) yielding a bounded-below functional, we proved the existence of minimizers for not too large a total charge $\operatorname{Tr}(\gamma)$. The functions $\beta(\nu)=\nu^m$ have themselves been shown to be related to certain Lieb-Thirring and Gagliardo-Nirenberg inequalities [23].

1.2 Multiconfiguration methods

We present in this section the multiconfiguration methods which aim in Chemistry at calculating an approximation of the Schrödinger eigenstates. It is a class of nonlinear models, parameterized by an integer K: when $K=N$ it is simply the Hartree-Fock model presented before, whereas when $K=+\infty$ one recovers the full linear Schrödinger theory. Our results presented in this section are contained in [46, 10, 49].

1.2.1 The multiconfiguration approach

The multiconfiguration approach is a natural generalization of the Hartree-Fock method in which one restricts the Schrödinger Hamiltonian to wavefunctions which are a linear combination of several Slater determinants instead of only one. This is justified by the remark that for any orthonormal basis $\{\psi_i\}$ of $L^2(\mathbb{R}^3,\mathbb{C})$, $\{\psi_{i_1}\wedge\cdots\wedge\psi_{i_N}\}_{i_1<\cdots<i_N}$ forms an orthonormal basis of the fermionic N-body space $\bigwedge_1^N L^2(\mathbb{R}^3,\mathbb{C})$.

An integer $K\geq N$ being fixed, we consider the subset of $\bigwedge_1^N L^2(\mathbb{R}^3)$ consisting of the wavefunctions Ψ which are *finite* linear combinations of the $\binom{K}{N}$ Slater determinants constructed from a set of K orthonormal functions $(\varphi_1,\cdots,\varphi_K)$ of $L^2(\mathbb{R}^3)$, i.e.

$$\Psi = \sum_{1\leq i_1<\cdots<i_N\leq K} c_{i_1\ldots i_N}\varphi_{i_1}\wedge\cdots\wedge\varphi_{i_N}. \tag{1.17}$$

In the multiconfiguration approach, both the coefficients $c_{i_1\ldots i_N}$ and the functions $(\varphi_1,\cdots,\varphi_K)$ (called *orbitals*) are variational parameters. When there is no ambiguity, we shall use the following notation

$$\Psi = \sum_{I\subset\{1,...,K\},\ |I|=N} c_I\ \Phi_I,$$

where $\Phi_I = \varphi_{i_1}\wedge\cdots\wedge\varphi_{i_N}$ when $I = \{i_1 < \cdots < i_N\}$. In Quantum Chemistry, multiconfiguration methods are usually called 'MCSCF' for *MultiConfiguration Self-Consistent Field* (the expression 'Self-Consistent Field' is often used in Chemistry to emphasize the nonlinearity of the model).

As the Hamiltonian $H^N(R, Z)$ is real, its eigenfunctions can be chosen real. This leads to the natural *assumption* that the coefficients c_I's as well as the orbitals φ_i's are both real. In the rest of this section, for the sake of simplicity we will always make this assumption, the results being very similar in the case of complex quantities. We use the simplified notation $L^2(\mathbb{R}^3) := L^2(\mathbb{R}^3, \mathbb{R})$, $H^1(\mathbb{R}^3) := H^1(\mathbb{R}^3, \mathbb{R})$, etc. Also, as the locations and charges (R, Z) of the nuclei will be fixed, we will often forget to mention them.

Following our purpose to describe the MCSCF approach, we introduce the manifold

$$\mathcal{M}^{N,K} = \left\{(c,\Phi)\in\mathbb{R}^{\binom{K}{N}}\times(H^1(\mathbb{R}^3))^K,\ \sum_{i_1<\cdots<i_N}|c_{i_1\dots i_N}|^2 = 1,\ \int_{\mathbb{R}^3}\varphi_i\varphi_j = \delta_{ij}\right\} \tag{1.18}$$

where we have used the notation

$$c = (c_{i_1\cdots i_N}) \in \mathbb{R}^{\binom{K}{N}}, \qquad \Phi = (\varphi_1, ..., \varphi_K) \in H^1(\mathbb{R}^3)^K$$

(we arrange the $c_{i_1\cdots i_N}$ in a column vector c using for instance the lexicographical order). The MCSCF energy functional that we denote here by $\mathcal{E}^{N,K}$, is defined by the formula

$$\mathcal{E}^{N,K}(c,\Phi) = \left\langle \Psi_{(c,\Phi)}, H^N(R,Z)\Psi_{(c,\Phi)}\right\rangle \tag{1.19}$$

$$\Psi_{(c,\Phi)} = \sum_{1\le i_1<\cdots<i_N\le K} c_{i_1\dots i_N}\varphi_{i_1}\wedge\cdots\wedge\varphi_{i_N},$$

and the MCSCF ground state energy then reads

$$E^{N,K}(R,Z) = \inf_{\mathcal{M}^{N,K}} \mathcal{E}^{N,K}. \tag{1.20}$$

An explicit expression of the functional $\mathcal{E}^{N,K}$ can be found in [46, Eq. (6)]. Let us point out that, whereas the Schrödinger energy functional $\Psi \mapsto \langle\Psi, H^N(R,Z)\Psi\rangle$ is quadratic, the MCSCF energy functional is not. Consequently, the MCSCF equations, namely the first order stationarity conditions for the critical points of $\mathcal{E}^{N,K}$ on the manifold $\mathcal{M}^{N,K}$, will be nonlinear. More precisely, $\mathcal{E}^{N,K}$ is not quadratic with respect to the orbitals φ_i's, but it is indeed quadratic with respect to the c_I's since

$$\mathcal{E}^{N,K}(c,\Phi) = \sum_{I,J} c_I c_J \left\langle \Phi_I, H^N(R,Z)\Phi_J\right\rangle = \sum_{I,J} c_I c_J (H_\Phi)_{IJ}$$

where (recall that $\Phi_I = \varphi_{i_1}\wedge\cdots\wedge\varphi_{i_N}$, when $I = \{i_1 < \cdots < i_N\}$)

$$(H_\Phi)_{IJ} = \left\langle \Phi_I, H^N(R,Z)\Phi_J\right\rangle. \tag{1.21}$$

In other words, H_Φ is the $\binom{K}{N}\times\binom{K}{N}$ matrix of the quadratic form associated with $H^N(R,Z)$ when it is restricted to the $\binom{K}{N}$-dimensional space $V_\Phi = \mathrm{Span}(\Phi_I)$. It can be seen that the *MCSCF equations* take the following general form [29, 46]

$$\begin{cases} \gamma_i\left(-\dfrac{\Delta}{2} + V_{R,Z}\right)\varphi_i + \displaystyle\sum_{1\le j,k,l\le K} b_{ijkl}\left((\varphi_j\varphi_k) * \frac{1}{|x|}\right)\varphi_l = \sum_{j=1}^K \lambda_{ij}\varphi_j,\ 1\le i\le K \\ H_\Phi\ c = \beta c, \end{cases} \tag{1.22}$$

where the b_{ijkl} are real numbers which can be expressed in terms of c. The first line of (1.22) is a system of K nonlinear coupled partial differential equations accounting for the stationarity conditions with respect to Φ; the symmetric matrix (λ_{ij}) is the Lagrange multiplier matrix associated with the orthonormality constraints on Φ. The numbers γ_i are called the *occupation numbers* and satisfy $0 \leq \gamma_i \leq 1$. They are the eigenvalues of the one-body density matrix γ_Ψ associated with the wavefunction Ψ (see [46] for details). A compact form of the first equations of (1.22) is given in [46]. The second equation is a simple eigenvalue problem and conveys the stationarity condition with respect to c.

Remark that in (1.17), all the Slater determinants that can be built with the functions φ_i are taken into account. Most often, this cannot be done in practice for $\binom{K}{N}$ is too large a number. It is then necessary to resort to an additional approximation consisting in dividing the electrons into two groups, the *inactive* electrons that are supposed to be correctly described by a Hartree-Fock type model, and the *active* electrons that mostly contribute to the correlation energy, and in using the MCSCF methodology for the active electrons only. This is the so-called Complete Active Space Self-Consistent Field (CASSCF) approach [70].

1.2.2 Existence and properties of minimizers

For algebraic reasons [3, 17, 58, 29, 46], there does not exist N-body wavefunctions of rank $K = N+1$ and when $N = 2$, all the Ψs have an even rank. A partial multiconfiguration method of rank $K = N+2$ was studied by Le Bris [43] who considered the minimization over doubly excited configurations

$$\Psi = \alpha\, \varphi_1 \wedge \cdots \wedge \varphi_N + \beta\, \varphi_1 \wedge \cdots \wedge \varphi_{N-2} \wedge \varphi_{N+1} \wedge \varphi_{N+2}. \tag{1.23}$$

He proved the existence of a minimum when $|Z| > N-1$ and showed the inequality $E^{N,N+2}(R,Z) < E^{N,N}(R,Z)$. The existence of a minimizer for $E^{N,K}(R,Z)$ for all $K \geq N$ and $N < |Z|+1$ was proved first by Friesecke in [29]. The result is the following:

Theorem 1.1 (Existence and properties of MCSCF ground state [29, 30, 44, 46]). *Let $K \geq N$ with $N < |Z|+1$. Then there exists a minimizer $(c,\Phi) \in \mathcal{M}_N^K$ for the variational problem $E^{N,K}(R,Z)$. Additionally, one has for every $K \geq N$,*

$$E^{N,K+2}(R,Z) < E^{N,K}(R,Z). \tag{1.24}$$

Lastly, any associated sequence of N-body wavefunction $\Psi_K = \sum_I c_I \Phi_I$ has a subsequence which converges strongly in $H^2((\mathbb{R}^3)^N)$ as $K \to \infty$ to a true ground state Ψ of the Schrödinger operator $H^N(R,Z)$.

To prove the existence of a minimizer for $E^{N,K}(R,Z)$, Friesecke used both concentration-compactness ideas of Lions [54, 55] and N-body geometric methods inspired by the HVZ theorem [35, 82, 89, 76, 69]. He proved the compactness of all minimizing sequences under the binding condition that

$$E^{N,K}(R,Z) < E^{N-1,K-1}(R,Z)$$

(compare with the HVZ criterion $E^N(R,Z) < \Sigma^N(R,Z) = E^{N-1}(R,Z)$ in the Schrödinger case), an inequality which can itself easily be proved using a test function when $N-1 < |Z|$.

Unfortunately, it is not obvious how to generalize this method to excited states and we gave a different proof of the existence of a ground state in [44, 46]. It was itself inspired of the proof given by Lions [56] for the Hartree-Fock model. The idea is to construct a specific Palais-Smale sequence with a second-order information (i.e. such that the second derivative is nonnegative up to a controlled error), using a generalization of Ekeland's lemma [26] due to Borwein and Preiss [7, 32]. Then, one shows that such a Palais-Smale sequence is indeed compact, using both the information that the first derivative goes to zero and that the second derivative is essentially nonnegative.

The convergence of the wavefunction as $K \to \infty$ was proved in [29]. The monotonicity property of the MCSCF energy (1.24) was shown independently in [30] and in [47], using ideas of Le Bris [43].

1.2.3 Nonlinear excited states

As we have mentioned before, the MCSCF theory is in Quantum Chemistry a method of choice for the calculation of *excited states*. However, contrarily to the ground state, it is not *a priori* clear how to define excited states for a nonlinear model such as MCSCF. As we will see, the energy functional $\mathcal{E}^{N,K}$ has many critical points on the manifold $\mathcal{M}^{N,K}$, many of them being probably purely due to the nonlinear nature of the model[1].

Taking into account the full nonlinearity of the problem and following ideas of Lions [56], a family of critical points of $\mathcal{E}^{N,K}$ on the manifold $\mathcal{M}^{N,K}$ satisfying very natural properties were constructed in [46]. The result is the following:

Theorem 1.2 (MCSCF Critical points [46]). *Let $K \geq N$ with $N < |Z| + 1$.*

(i) (Existence of infinitely many critical points). *There exists an infinite sequence $\{c_i, \Phi_i\}_{i\geq 1} \subset \mathcal{M}^{N,K}$ of critical points of the functional $\mathcal{E}^{N,K}$ on the manifold $\mathcal{M}^{N,K}$, satisfying $\mathcal{E}^{N,K}(c_i, \Phi_i) < E^0(R,Z)$ for all $i \geq 1$ and*

$$\lim_{i\to\infty} \mathcal{E}^{N,K}(c_i, \Phi_i) = E^0(R,Z). \tag{1.25}$$

In addition, the corresponding sequence $\{\Psi_i\}_{i\geq 1}$ of N-body wavefunctions satisfies

$$\lim_{i\to\infty} \|\nabla\Psi_i\|_{L^2} = 0.$$

(ii) (Existence of finitely many nonlinear excited states). *There exists $\binom{K}{N}$ critical points (c_d^K, Φ_d^K), $0 \leq d \leq \binom{K}{N} - 1$, of the functional $\mathcal{E}^{N,K}$ on the manifold $\mathcal{M}^{N,K}$, satisfying*

$$\lambda_d^N(R,Z) \leq \mathcal{E}^{N,K}(c_d^K, \Phi_d^K) := \lambda_d^{N,K} \tag{1.26}$$

and

$$\lim_{K\to\infty} \mathcal{E}^{N,K}(c_d^K, \Phi_d^K) = \lambda_d^N(R,Z).$$

As announced previously the first part (i) shows that the functional $\mathcal{E}^{N,K}$ has infinitely many critical points on the manyfold $\mathcal{M}^{N,K}$, for any fixed K. The so-constructed points are the natural generalization of the ones which were introduced by Lions in [56] in the Hartree-Fock case. Note that (1.25) together with the fact that $\lambda_d^N(R,Z) \to \Sigma^N(R,Z) = E^{N-1}(R,Z) < E^0(R,Z)$ when $N > 1$ and as $d \to \infty$, imply that the critical points of Part (i) are probably not very much related to eigenfunctions of $H^N(R,Z)$. We do not call them "excited states".

Some critical points which behave as expected in the limit $K \to \infty$ are constructed in Part (ii). We believe that the sequence of wavefunctions $\{\Psi_d^K\}$ corresponding to (c_d^K, Φ_d^K) converges strongly in $H^2(\mathbb{R}^{3N})$ (up to a subsequence) as $K \to \infty$ to a dth excited state Ψ_d solution of $H^N(R,Z)\Psi_d = \lambda_d^N(R,Z)\Psi_d$, but we could not prove it.

The method of proof of Theorem 1.2 is very much inspired from the one of Lions in the Hartree-Fock case [56, 32]. A certain minimax method is defined and again a special Palais-Smale sequence including a second-order Morse-type information is considered. This time, one uses a result of Fang and Ghoussoub [28] to get such a sequence. The strong convergence of the Palais-Smale sequence when $N < |Z| + 1$ is obtained like in the minimization case.

There is a natural definition for the MCSCF excited state energies, which is indeed the one which is mostly used in Quantum Chemistry (see, e.g. [73, 84, 85] and the references of [10]). Let us denote by $\tilde{\lambda}_d^{N,K}(\Phi)$, $d = 0, ..., \binom{K}{N} - 1$, the $\binom{K}{N}$ eigenvalues of the Hamiltonian matrix H_Φ defined in (1.21), depending on the orbitals $\Phi = (\varphi_1, ..., \varphi_K)$. By the usual Rayleigh-Ritz formula, one deduces that

$$\lambda_d^N(R,Z) \leq \tilde{\lambda}_d^{N,K}(\Phi).$$

---

[1] Contrarily to what is sometimes seen in the mathematical literature, we do *not* use the expression "excited state" for every critical points of our functional. We will reserve this name to the critical points which have some relation with the true excited states of the Schrödinger model, in the sense detailed above.

Sch.	$K = +\infty$	$\lambda_0^N(R, Z)$	$\lambda_1^N(R, Z)$	$\cdots$	$\cdots$	$\cdots$	$\lambda_d(R, Z)$	$\cdots$	$\cdots$
MC	$\vdots$	$\uparrow$	$\uparrow$				$\uparrow$		
	K	$\lambda_0^{N,K}$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\lambda_{\binom{K}{N}-1}^{N,K}$	
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$				
	$K = N + 2$	$\lambda_0^{N,N+2}$	$\lambda_1^{N,N+2}$	$\cdots$	$\lambda_{\binom{N+2}{N}-1}^{N,N+2}$				
HF	$K = N$	$\lambda_0^{N,N}$							

Table 1.1: The nonlinear excited states constructed in Part (*ii*) of Theorem 1.2.

It is therefore natural to define the following variational method:

$$\tilde{\lambda}_d^{N,K} = \inf_{\Phi} \mu_d^K(\Phi), \tag{1.27}$$

that is to say, quoting [73], *"the MCSCF energy results from minimizing the appropriate eigenvalue of the Hamiltonian matrix with respect to orbital variations"*.[2] It can actually be proved that $\tilde{\lambda}_d^{N,K} \searrow \lambda_d^N(R, Z)$ as $K \to \infty$.

However, (1.27) is a minimization of an eigenvalue of a symmetric matrix depending on a parameter Φ. This type of variational method is generally very ill-posed mathematically (even in finite dimensions) and it is not at all guaranteed that there will be a critical point at the level $\tilde{\lambda}_d^{N,K}$ (see [46, 10, 49] for comments in this direction). Indeed we believe that most of the convergence problems encountered in practical computations (like the so-called *root flipping* [86, 85, 84]) are due to this issue.

The problems raised by Definition (1.27) have already been described and studied in details in the Chemistry literature [33, 36, 37, 66] by the team of the DALTON software [1]. They proposed a different definition of excited states by requiring that a d^{th} excited state is a stationary state with Morse index at most d. Such states are computed in DALTON by a well-behaved Newton-type algorithm followed by a trust region method, which does not lead to any root-flipping problem. However, it was not proposed how to get the correct critical points. The minimax method used to construct the ones of Theorem 1.2 fills this gap.

We now quickly explain how the min-max variational methods are defined. If a group G acts on two topological spaces X and Y, we recall that a function $\varphi : X \to Y$ is called *G-equivariant* if $\varphi(g \cdot x) = g \cdot \varphi(x)$ for all $g \in G$ and $x \in X$. We denote by $\mathcal{C}_G(X, Y)$ the set of all continuous G-equivariant functions. In our case, we consider, as in [56, 32, 68], min-maxing methods of the form:

$$\min_{f \in \mathcal{C}_G(S^d, \mathcal{M}^{N,K})} \max_{(c,\Phi) \in f(S^d)} \mathcal{E}^{N,K}(c, \Phi) \tag{1.28}$$

where $G = \mathbb{Z}_2 \simeq \{\pm 1\}$ acts obviously on the Euclidian sphere S^d of $\mathbb{R}^{d+1}$. The action of $\mathbb{Z}_2$ on $\mathcal{M}^{N,K}$ is defined using the fact that $\mathcal{E}^{N,K}$ is even in c and Φ:

$$\mathcal{E}^{N,K}(c, \Phi) = \mathcal{E}^{N,K}(c, -\Phi) = \mathcal{E}^{N,K}(-c, \Phi).$$

Hence two natural group actions of $\mathbb{Z}_2$ may be considered on $\mathcal{M}^{N,K}$:

$$(-) \cdot_{\Phi} (c, \Phi) = (c, -\Phi), \tag{1.29}$$

$$(-) \cdot_{c} (c, \Phi) = (-c, \Phi). \tag{1.30}$$

---

[2] Note that in Chemistry, it is always assumed that the model contains at least $d + 1$ configurations in order to compute an approximation of the dth excited state. For instance no excited state is defined in the Hartree-Fock model.

A function $f : S^d \longrightarrow \mathcal{M}^{N,K}$ is said to be $(\mathbb{Z}_2)_c$-equivariant when $f(x) = (c, \Phi) \implies f(-x) = (-c, \Phi)$ and $(\mathbb{Z}_2)_\Phi$-equivariant when $f(x) = (c, \Phi) \implies f(-x) = (c, -\Phi)$. Let us denote by p the projection on the MCSCF coefficients: $p(c, \Phi) = c$. Then we note that when f is a continuous $(\mathbb{Z}_2)_c$-equivariant function, $p \circ f : S^d \to S^{\binom{K}{N}-1}$ is a continuous and odd function. By the Borsuk-Ulam Theorem, such functions only exist when $d + 1 \leq \binom{K}{N}$. This leads us to the following definitions:

$$\mu_d^{N,K} = \inf_{f \in \mathcal{C}_{(\mathbb{Z}_2)_\Phi}(S^d, \mathcal{M}^{N,K})} \max_{(c,\Phi) \in f(S^d)} \mathcal{E}^{N,K}(c, \Phi) \tag{1.31}$$

for $d \geq 1$, and

$$\lambda_d^{N,K} = \inf_{f \in \mathcal{C}_{(\mathbb{Z}_2)_c}(S^d, \mathcal{M}^{N,K})} \max_{(c,\Phi) \in f(S^d)} \mathcal{E}^{N,K}(c, \Phi) \tag{1.32}$$

for $0 \leq d \leq \binom{K}{N} - 1$.

The critical levels $\mu_d^{N,K}$ and $\lambda_d^{N,K}$ respectively furnish the critical points of Part (i) and (ii) in Theorem 1.2. Let us introduce

$$\begin{array}{rccc} \Pi^{N,K}: & \mathcal{M}^{N,K} & \longrightarrow & \bigwedge_1^N H^1(\mathbb{R}^3) \\ & (c, \Phi) & \longmapsto & \Psi = \sum_I c_I \Phi_I, \end{array}$$

the natural projection from the one-body space into the N-body space. It is easy to see that the $(d+1)$st eigenvalue can also be obtained through a nonlinear minimax principle:

$$\lambda_d^N(R, Z) = \min_{g \in \Theta_d} \max_{\Psi \in g(S^d)} \langle \mathcal{H}\Psi, \Psi \rangle$$

where Θ_d is the collection of all odd continuous map from S^d into the sphere of $\bigwedge_1^N H^1(\mathbb{R}^3)$. Note that the group action defined in (1.30) was chosen in such a way that for all $(\mathbb{Z}_2)_c$-equivariant function $f : S^d \to \mathcal{M}_N^K$, it holds $g = \Pi_N^K \circ f \in \Theta_d$. Hence we deduce that $\lambda_d^{N,K} \geq \lambda_d^N(R, Z)$. On the other hand, we have for all fixed Φ

$$\lambda_d^{N,K} \leq \inf_{\substack{f \in C^0\left(S^d, S^{\binom{K}{N}-1}\right) \\ f \text{ odd}}} \max_{(c,\Phi) \in f(S^d) \times \{\Phi\}} \mathcal{E}^{N,K}(c, \Phi) = \tilde{\lambda}_d^{N,K}(\Phi)$$

hence we also deduce that $\lambda_d^{N,K} \leq \tilde{\lambda}_d^{N,K}$. Therefore we see that our definition (1.32) is the natural extension of (1.27) taking into account the full nonlinearity of the problem.

1.2.4 A new algorithm for the computation of the first excited state

A new computational method based on the definition (1.32) was proposed and studied for the first excited state in a collaboration with Éric Cancès and Hervé Galicher in [10] (see also the review [49]). Using the fact that $\mathcal{E}^{N,K}$ is even with respect to c, it can be seen that the minimax level (1.32) for $d = 1$ may be rewritten

$$\lambda_1^{N,K} = \inf_{(c,\Phi) \in \mathcal{M}^{N,K}} \left\{ \inf_{\gamma \in \Gamma_{(c,\Phi)}} \sup_{t \in [0;1]} \mathcal{E}^{N,K}(\gamma(t)) \right\} \tag{1.33}$$

where

$$\Gamma_{(c,\Phi)} = \left\{ \gamma \in C^0\left([0;1], \mathcal{M}^{N,K}\right),\ \gamma(0) = (c, \Phi),\ \gamma(1) = (-c, \Phi) \right\}.$$

Notice that the inf − sup problem which is in brackets in (1.33) is a usual mountain-pass problem between (c, Φ) and $(-c, \Phi)$. We now *conjecture* that when K is large enough, a global minimizer of the MCSCF energy (c_0^K, Φ_0^K) as provided by Theorem 1.1 is also a minimizer of the outer minimization in (1.33). Therefore, we are able to simplify the resolution of problem (1.33) as follows: we clamp both ends of the trial paths at (c_0^K, Φ_0^K) and $(-c_0^K, \Phi_0^K)$ respectively, and solve the mountain pass problem

$$\lambda_1^{N,K} = \inf_{\substack{\gamma \in C^0\left([0;1], \mathcal{M}^{N,K}\right), \\ \gamma(0)=(c_0^K, \Phi_0^K),\ \gamma(1)=(-c_0^K, \Phi_0^K)}} \sup_{t \in [0;1]} \mathcal{E}_N^K(\gamma(t)). \tag{1.34}$$

Notice that (1.34) mimics the fact that, in the linear case, one can obtain the second eigenfunction Ψ_1 of $H^N(R,Z)$, as a mountain pass point between Ψ_0 and $-\Psi_0$, where $H^N(R,Z)\Psi_0 = E^N(R,Z)\Psi_0$, see Fig. 1.2.

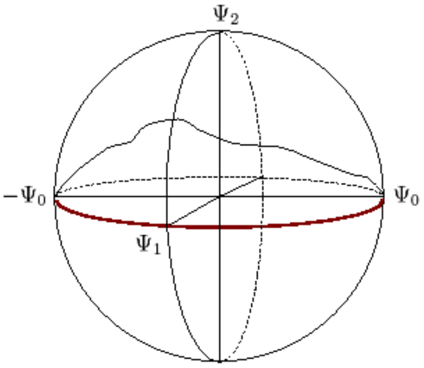


Figure 1.2: The true (Schrödinger) first excited state Ψ_1 can be obtained as a mountain pass point between the two ground states $\pm\Psi_0$: any path linking Ψ_0 and $-\Psi_0$ necessarily intersects the orthogonal of $\text{Span}\Psi_0$, on which the energy is $\geq \lambda_1$.

In practice, solving a mountain-pass problem is rather demanding in terms of CPU time since one has to deform paths. Therefore, a not too tight convergence criteria was chosen in [10] to stop the path optimization step. The state of highest energy on the final path is then used as initial guess in a Newton-like procedure to solve (1.22).

We have found many algorithms in the literature for the optimization of paths (often applied to the simulation of chemical reactions on potential energy surfaces) [72, 67, 34, 14, 27, 21, 25], some of them being quite peculiar in our opinion. The method that was proposed in [10] for the deformation of paths, and which seems to give good results on our problem, is of general concern and could therefore also be useful for some other problems. It was partially inspired of [15, 16]. The idea is to sample a path by a certain number of points on the corresponding manifold, and use the gradient flow of the functional to be optimized to deform the path. A naive procedure consists in just moving each point of the path in the direction opposite to the gradient, but in this case the points will fall down in one of the valley. Many methods have been introduced to circumvent this issue. It was for instance proposed to attach each point to its neighbor by means of a string, or to use the component of the gradient normal to the path to move the points. A better approach that we have used in [10] consists instead of computing at each step new points on the paths which are better distributed in some sense, hence avoiding them to fall into the valleys. Details can be found in [10].

We have applied the above strategy to the case of two-electron systems. We considered the singlet state for which the antisymmetry of the wavefunction is encoded in the spin variable, hence the spacial component is symmetric, which helped us in reducing the complexity of the calculation.

It is well-known in Chemistry that the true Schrödinger eigenfunctions have special symmetries. The ground state is an even function, $\psi_0(-x,-y) = \psi_0(x,y)$, whereas the first excited state is an odd function, $\psi_1(-x,-y) = -\psi_1(x,y)$. Hence the first excited state is indeed a ground state of its symmetry subspace: Chemists would usually impose this symmetry and minimize. To emphasize

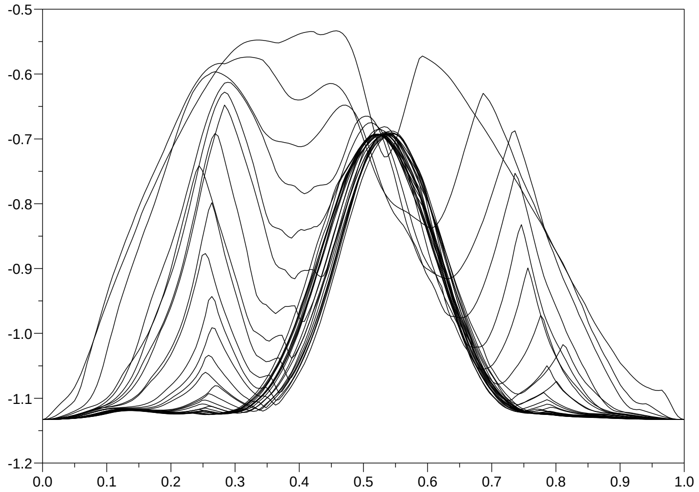


Figure 1.3: Energy along the successive deformed paths generated by the algorithm of [10] for the computation of the first singlet excited state of H_2, with $R = 1$ Å.

the possible difficulties when dealing with MCSCF methods, we have not imposed any symmetry (except the one coming from the spin) on our wavefunctions.

We used an MCSCF method with 4 orbitals, i.e. the wavefunction takes the form

$$\psi(x, y) = \sum_{i=1}^{4} c_i \varphi_i(x)\varphi_i(y),$$

each φ_i being expanded in a selected one-body basis. We treated both Helium-like atoms and the Hydrogen molecule using the cc-pVDZ basis whose size is 10. The MCSCF ground state was always found to be an even function, as expected. Figure 1.3 shows the successive paths created by our algorithm for the computation of the first singlet excited state of the H_2 molecule with an interatomic distance of 1 Å.

The computed value of the vectors $(c_i)_{i=1}^4$ in both the ground state and the excited state is as follows:

$$c_{\min} = \begin{bmatrix} 0.9860929 \\ -0.1564182 \\ -0.0548179 \\ -0.0122131 \end{bmatrix} \qquad c_{1\text{ex}} = \begin{bmatrix} -0.7086355 \\ 0.7051798 \\ 0.0166917 \\ 0.0166917 \end{bmatrix}.$$

As is well-known by chemists, one finds that the ground state has one dominating coefficient, showing that the Hartree-Fock approximation would yield a quite good result. However the first excited state has two c_i's of similar size, hence the corresponding wavefunction would be poorely described by a Hartree-Fock state. What was surprising to us was that the computed nonlinear first excited state wavefunction was found to be *almost* odd, but not quite. Hence there is a symmetry breaking phenomenon occurring: due to the nonlinearity, the MCSCF wavefunction does not exactly share the symmetry of the true first excited state. It is only in the limit of infinitely many Slater determinants that the first nonlinear excited state will have the correct symmetry.

Also in [10], the problem (1.27) corresponding to minimizing the second eigenvalue of the Hamiltonian matrix was considered. A good candidate for a solution was found. The corresponding wavefunction is even, showing that when no symmetry is imposed on the wavefunction, one indeed has $\lambda_d^{N,K} < \tilde{\lambda}_d^{N,K}$ (for the corresponding discretized problems). This was already mentioned earlier in [64].

1.3 A new reduced density matrix approach

Recently, there has been a regain of interest in the so-called *2-body reduced density matrix method.* This method allows to calculate an approximation of the ground state energy $E^N(R,Z)$ (it does not allow to calculate excited states). We present this method very quickly here, and refer to the article [13] written in collaboration with É. Cancès and G. Stoltz for a more precise presentation (or, even better, to the PhD thesis of G. Stoltz [79] which contains many more details).

The first step consists in noting that

$$E^N(R,Z) = \inf_{\substack{\Upsilon\in\mathcal{B}\left(\bigwedge_1^N L^2(\mathbb{R}^3)\right)\\ \Upsilon\geq 0,\ \mathrm{Tr}(\Upsilon)=1}} \mathrm{Tr}\left(H^N(R,Z)\Upsilon\right)$$

which is interpreted by saying that for a linear model, the ground state energy among mixed states is the same as the one among pure states. As early as in 1951, it was noticed by Coleman that the above electronic N-body ground-state energy could be expressed using only the two-body reduced density matrices (2-RDM), defined as $\Gamma := N(N-1)\mathrm{Tr}_{3\cdots N}\Upsilon$, where $\mathrm{Tr}_{3\cdots N}$ denotes the partial trace in the last $N-2$ variables. This can be expressed in terms of the kernels of the operators as

$$\Gamma(x_1,x_2;x_1',x_2') := N(N-1)\int_{\mathbb{R}^3}\cdot\int_{\mathbb{R}^3}\Upsilon(x_1,x_2,x_3,...,x_N;x_1',x_2',x_3,...,x_N)dx_3\cdots dx_N. \quad (1.35)$$

It can then be verified that for every self-adjoint operator Υ,

$$\mathrm{Tr}_{\bigwedge_1^N L^2(\mathbb{R}^3)}(H^N(R,Z)\Upsilon) = \mathrm{Tr}_{L^2(\mathbb{R}^3)\wedge L^2(\mathbb{R}^3)}(K_N(R,Z)\Gamma)$$

where $K_N(R,Z)$ is a two-body operator defined as

$$K_N(R,Z) = \frac{h_{x_1}+h_{x_2}}{2(N-1)} + \frac{1}{2|x_1-x_2|} + E^0(R,Z),$$

with $h = -\Delta/2 + V_{R,Z}$. Hence

$$E^N(R,Z) = \inf_{\substack{\Gamma\in\mathcal{C}_N\\ \mathrm{Tr}(\Gamma)=N(N-1)}} \mathrm{Tr}\left(K_N(R,Z)\Gamma\right) \quad (1.36)$$

where $\mathcal{C}_N$ denotes the *cone of N-representable two-body density matrices*

$$\mathcal{C}_N := \{\Gamma \mid \text{ there exists } \Upsilon\geq 0 \text{ such that (1.35) holds}\}.$$

At a conference in 1959, Coulson proposed to completely eliminate wavefunctions from Quantum Chemistry, since all the electronic ground-state properties of molecular systems can be computed from the 2-RDM [20, 58, 59]. Unfortunately, the set $\mathcal{C}_N$ of N-representable 2-RDM is not known explicitly. Some mathematical characterizations were provided [41, 18, 19] but they could not be used to derive a numerical method with a complexity of a lower order than the usual N-body problem. It was shown in [57] that the question to know whether a density matrix is N representable or not, is quantum Merlin-Arthur complete (this is the quantum generalization of nondeterministic polynomial time (NP) complete), hence NP hard.

Recently a new interest in the Reduced Density Matrix (RDM) approach arose. Impressive numerical results have been obtained by two different algorithms for semidefinite programming: primal-dual interior point methods [65, 60, 88, 31], or an augmented Lagrangian formulation using matrix factorization of the 2-RDM [62, 61, 63]. These results use a small number of known *necessary conditions* of N-representability. All these conditions take the very general form

$$\mathcal{L}_i(\Gamma) \geq 0$$

for some linear operators $\mathcal{L}_i : S\left(L^2(\mathbb{R}^3) \wedge L^2(\mathbb{R}^3)\right) \to S(\mathfrak{H}_i)$ and some Hilbert space $\mathfrak{H}_i$. The simpler methods use three conditions denoted as P, Q and G, whereas more complicated methods use two additional conditions T_1 and T_2. The P, Q and G conditions may be obtained in a second-quantized formalism by writting that any state must satisfy $\omega(c^\dagger c) \geq 0$, with, respectively, $c = a_i^\dagger a_j^\dagger$, $c = a_i a_j$ or $c = a_i^\dagger a_j$. One can obtain other conditions by considering all possible products of 3 creation and annihilation operators, but they of course involve the three-body density matrix. The T_1 and T_2 conditions are obtained by finding an appropriate linear combination of the 3-body inequalities, in order to get conditions where only the two-body density matrix appears.

The unknown cone $\mathcal{C}_N$ is therefore replaced by a bigger cone

$$\mathcal{C}_N^{\text{app}} := \{\Gamma \mid \mathcal{L}_i(\Gamma) \geq 0 \text{ for } i = 1, ..., \ell\}$$

and the associated ground state energy is a *lower bound* to the true energy $E^N(R, Z)$, defined as

$$E_{\text{app}}^N(R, Z) = \inf_{\substack{\Gamma \in \mathcal{C}_N^{\text{app}} \\ \text{Tr}(\Gamma) = N(N-1)}} \text{Tr}\left(K_N(R, Z)\Gamma\right). \tag{1.37}$$

In [13] we have noticed that by duality (1.37) may be written

$$E_{\text{app}}^N(R, Z) = N(N-1) \sup\{\mu \mid K_N(R, Z) - \mu \in (\mathcal{C}_N^{\text{app}})^*\} \tag{1.38}$$

where $(\mathcal{C}_N^{\text{app}})^*$ is the cone dual to $\mathcal{C}_N^{\text{app}}$, which is simply

$$(\mathcal{C}_N^{\text{app}})^* = \left\{ \sum_{i=1}^{\ell} (\mathcal{L}_i)^* B_i \ : \ B_i \in S(\mathfrak{H}_i), \ B_i \geq 0 \right\}.$$

The advantage of our formulation (1.38) is that it is a one-dimensional problem. Let us introduce the distance to the dual cone $(\mathcal{C}_N^{\text{app}})^*$:

$$\delta(\mu) = \text{dist}\left(K_N(R, Z) - \mu, (\mathcal{C}_N^{\text{app}})^*\right).$$

Denoting $\mu_{\text{app}}^* = E_{\text{app}}^N/(N(N-1))$, it can be seen that $\delta \equiv 0$ on $(-\infty, \mu_{\text{app}}^*]$, that δ is increasing on $[\mu_{\text{app}}^*, \infty)$ and convex on $\mathbb{R}$. Also δ^2 is continuously differentiable on $\mathbb{R}$, thus δ is continuously differentiable on $\mathbb{R} \setminus \{\mu_{\text{app}}^*\}$. Its derivative can indeed be explicitely computed: $\delta'(\mu) = -\text{Tr}(K_N - \mu - A_\mu) \left\|K_N - \mu - A_\mu\right\|^{-1}$ where A_μ denotes the projection of $K_N - \mu$ onto the polar cone $(\mathcal{C}_N^{\text{app}})^*$. To illustrate the above properties, a plot of $\delta(\mu)$ for N_2 in a STO-6G basis set is provided in Figure 1.4.

In order to compute μ_{app}^*, we use a Newton-like scheme that strongly exploits the above mentioned properties in a natural way: starting from an initial energy above μ_{app}^* (such as the Hartree-Fock energy for instance) and using the convexity of the function δ, the Newton algorithm ensures that the energy μ decreases at each step of the optimization process and converges to μ_{app}^*. The right derivative of δ at μ_{app}^* being always positive, the convergence rate is guaranteed to be at least superlinear. Of course, the most difficult part of the algorithm is the computation of the distance $\delta(\mu)$ to the cone, and of the projection A_μ of $K_N - \mu$. To this end, we chose to minimize, for a given μ, the objective function

$$J_\mu(B) = \frac{1}{2} \left\| K_N - \mu - \sum_{i=1}^{\ell} (\mathcal{L}_i)^* B_i \right\|^2,$$

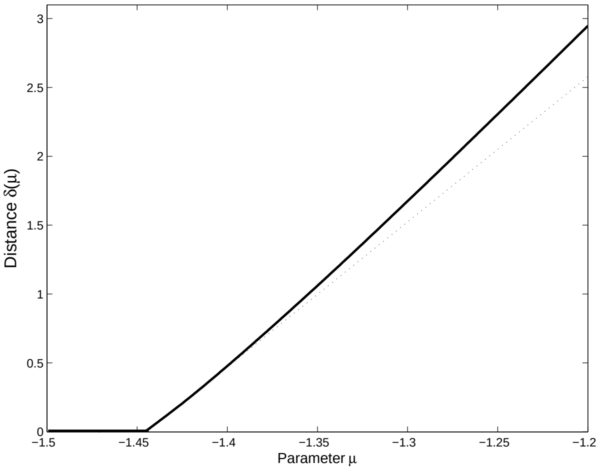


Figure 1.4: Distance of $K_N(R,Z) - \mu$ to the dual cone $(\mathcal{C}_N^{\rm app})^*$ as a function of μ for N_2 in a STO-6G basis set [13]. The tangent at the estimated value for $\mu_{\rm app}^* \simeq -1.4457$ is also displayed (dotted line). The Full CI and Hartree-Fock values are, in the same basis, $\mu_{\rm CI} = -1.4454$ and $\mu_{\rm HF} = -1.4435$.

under the constraints $B_i \geq 0$ $(i = 1...\ell)$. The above minimization is performed using a classical limited-memory BFGS algorithm [6]. Computing $\delta(\mu)$ with sufficient accuracy when μ is close to $\mu_{\rm app}^*$ can be difficult because the minimization of $J_\mu(B)$ then is ill-conditioned. We therefore consider a "truncated" version of the Newton algorithm where μ is updated by a fraction $0 < a \leq 1$ of the Newton step. We then use the linearity of δ for values close to $\mu_{\rm app}^*$ to devise a stopping criterion limiting the number of iterations.

Several numerical results using this algorithm are provided in [13]. In general, we have observed that the function δ is almost linear in quite large a right neighborhood of $\mu_{\rm app}^*$ (see Figure 1.4). One iteration of the Newton algorithm already provides a very correct approximation of the exact RDM energy, even when starting from the Hartree-Fock level. Usually, only 3 or 4 Newton iterations are necessary to achieve convergence.

1.4 A mountain pass for reacting molecules

1.4.1 Adiabatic reactions

In the previous sections we have been concerned with the approximation of the eigenvalues $\lambda_d^N(R,Z)$ of the Schrödinger Hamiltonian $H^N(R,Z)$ for fixed positions $R = (R_m)$ of the nuclei. One of the main goals of Chemistry is precisely to understand the behavior of the system when the nuclei can move (chemical reactions). In principle one should consider a coupled time-dependent system but one approximation which is very often done by chemists is to plot (an approximation of) the functions $R \mapsto \lambda_d^N(R,Z)$ and, assuming the eigenvalues do not cross, state that understanding the structure of these surfaces yields an information on the behavior of the system. We will not address here the issue of justifying the previous claim and we refer for instance to [80].

It was proved by Lieb and Thirring in [53] that for neutral molecules the function $R \mapsto \lambda_0^N(R,Z)$ always has at least one minimum. Here we are interested in the case for which $R \mapsto \lambda_0^N(R,Z)$ has two local minima, corresponding to two locally stable positions of the nuclei. Such situations are very common in Chemistry. An example is given by the molecule HCN which also has a stable isomer of the form CNH, see Fig. 1.5. In this case, it is natural to consider the usual mountain pass variational problem between the two local minima. A path linking these two minima is usually

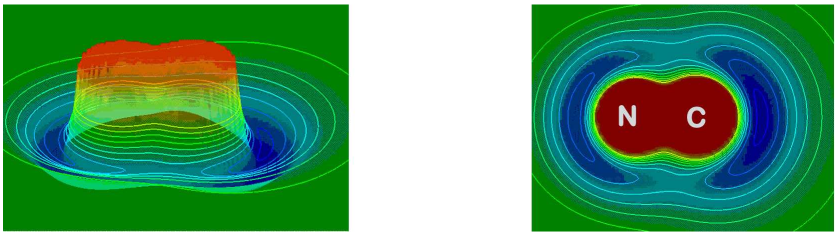


Figure 1.5: The Potential Energy Surface of HCN with C and N fixed (i.e. the value of the first Schrödinger eigenvalue when varying the position of H in the plane containing the three atoms), calculated in [81] using an MCSCF method.

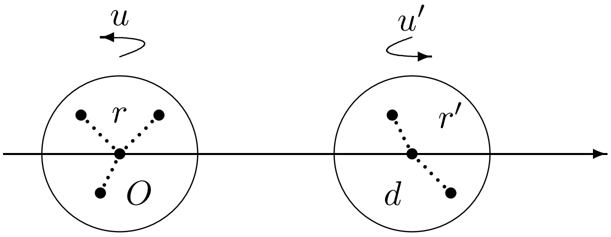


Figure 1.6: The system composed of two molecules studied in [45, 48].

interpreted as an adiabatic (infinitely slow) chemical reaction. Computing a good approximation of an optimal path as well as of the mountain pass is very important for chemists.

In this section we present some results of [45, 48]. For the sake of simplicity, we consider a molecular system composed of only two subsystems which are free to move and refer to [45] for a more general setting. This means we define, see Fig. 1.6,

$$R(d,u,u') = (u\cdot r, d\vec{v} + u'\cdot r'), \qquad Z = (z,z'),$$

where $r = (0, r_2, ...r_m) \in (\mathbb{R}^3)^m$, $r' = (0, r'_2, ..., r'_{m'}) \in (\mathbb{R}^3)^{m'}$ and $(z,z') \in (\mathbb{R}^+)^m \times (\mathbb{R}^+)^{m'}$ are fixed. The three variables are the rotations of the sub-molecules $u, u' \in SO(3)$ and the distance between them $d \in \mathbb{R}\setminus\{0\}$. We assume that the global molecule is neutral, i.e. $|Z| = |z| + |z'| = N$. For simplicity, we also define

$$H^N(d,u,u') = H^N(R(d,u,u'), Z), \qquad E^N(d,u,u') := \inf\sigma\{H^N(d,u,u')\}.$$

and

$$\mathcal{E}(d,u,u',\Psi) := \left\langle H^N(d,u,u')\Psi, \Psi\right\rangle$$

for all $\Psi \in \bigwedge_1^N H^1(\mathbb{R}^3,\mathbb{C})$, normalized in L^2.

We assume that $(d,u,u',\Psi) \mapsto \mathcal{E}(d,u,u',\Psi)$ possesses two local minima $M_1 = (d_1,u_1,u'_1,\Psi_1)$ and $M_2 = (d_2,u_2,u'_2,\Psi_2)$. Up to a rotation, we can always suppose that $d_1, d_2 > 0$. Next we define the mountain pass level as

$$\boxed{c = \inf_{\substack{\gamma\in C^0([0,1]\,;\,\mathcal{M}),\\ \gamma(0)=M_1,\ \gamma(1)=M_2}} \max_{t\in[0,1]} \mathcal{E}(\gamma(t))} \tag{1.39}$$

where

$$\mathcal{M} := (0,\infty)\times SO(3)^2 \times \left\{\Psi \in \bigwedge_1^N H^1(\mathbb{R}^3,\mathbb{C}) \ |\ \|\Psi\|_{L^2} = 1\right\}.$$

We believe that the following conjecture holds true:

Conjecture 1.1. *There always exists a critical point $M = (d, u, u', \Psi)$ of the energy $\mathcal{E}$, of Morse index at most one, at the level c.*

In [45, 48], the conjecture was proved under some assumptions on the "molecules at infinity", as will be explained below. We note that the proof of Lieb-Thirring [53] cannot be directly applied to this problem: it is not sufficient to know that at infinity there is one (unknown) orientation which yields an attractive interaction. If we want to prove that it is not energetically favorable for a min-maxing sequence of paths to escape to infinity, we also need have some information on the directions along which the energy decreases "at infinity" (for instance via the negative eigenspaces of the Hessian).

Let us remark that we use the wavefunction as variable in our minimax method. This is because of some regularity problem in the case for which the first eigenvalue is degenerated at the level c. Using the fact that our wavefunctions takes complex values, it was indeed shown in [45] that

$$c = \inf_{\substack{R\in C^0\big([0,1]\ ;\ (0,\infty)\times SO(3)^2\big),\\ R(0)=(d_1,u_1,u_1'),\ R(1)=(d_2,u_2,u_2')}} \max_{t\in[0,1]} E^N(R(t)).$$

The same holds for real functions, but maybe only for a correct choice of Ψ_1 and Ψ_2.

1.4.2 Loss of compactness

As usual we have to understand the behavior of Palais-Smale sequences in the case for which $d \to \infty$ on a sequence of paths approximating (1.39), i.e. when there is a loss of compactness. The following was proved in [45]:

Theorem 1.3 (Description of possible loss of compactness [45])**.** *We have the following alternative:*

- ***either** one can find a Palais-Smale sequence $\{d_n, u_n, u_n', \Psi_n\} \in \mathcal{M}$ with $\{d_n\}$ bounded and which converges to a critical point (d, u, u', Ψ) of $\mathcal{E}$ at the level c, with a Morse index at most one and such that*

  $$c = E^N(d, u, u'), \qquad H^N(d, u, u')\Psi = c\Psi;$$

- ***or** there exists $\delta_n \to \infty$ and $\gamma_n \in C^0([0,1], \mathcal{M})$ with $\gamma(0) = M_1$ and $\gamma(1) = M_2$, such that*

  $$\lim_{n\to\infty} \max_{t\in[0,1]} \mathcal{E}^N(\gamma_n(t)) = c \ \text{ and } \ d_n(t) \leq \delta_n \Longrightarrow \mathcal{E}^N(\gamma_n(t)) < c.$$

  In this case, one has

  $$c = \min\big\{E^{n_1}(r, z) + E^{n_2}(r', z') \mid n_1 + n_2 = N\big\}. \tag{1.40}$$

The proof uses a duality theory developed by Ghoussoub [32]. One can also study the sequence $\{\Psi_n\}$ "at the top of γ_n" and show that it converges to a "critical point at infinity" [45].

Of course, understanding the "molecules at infinity", i.e. the ground states of the Hamiltonians $H^{n_1}(r, z)$ and $H^{n_1}(r, z)$ with $n_1 + n_2 = N$ minimizing (1.40) will be extremely useful. We mention the following simple result which is contained in [45]:

Proposition 1.1. *Assume that $(N_1, N_2) \in \mathbb{N}^2$ is such that*

$$E^{N_1}(r, z) + E^{N_2}(r', z') = \min\{E^{n_1}(r, z) + E^{n_2}(r', z') \mid n_1 + n_2 = N\}.$$

Then $E^{N_1}(r, z) < E^{N_1-1}(r, z)$ and $E^{N_2}(r, z) < E^{N_2-1}(r, z)$. Thus both $H^{N_1}(r, z)$ and $H^{N_2}(r', z')$ have a ground state.

We provide the proof which is a very simple application of the properties of the point spectrum of Coulomb Hamiltonians depending on the number of electrons.

Proof. If for instance $E^{N_1}(r,z) = E^{N_1-1}(r,z)$, then $N_1 \geq |z|+1$ by Zhislin's Theorem [90, 91]. Thus $N_2+1 \leq |z'|$ since $N_1+N_2 = |z|+|z'|$. We deduce that $E^{N_2+1}(r',z') < \Sigma^{N_2+1}(r',z') = E^{N_2}(r',z')$ by the HVZ Theorem. But then this implies

$$E^{N_1-1}(r,z) + E^{N_2+1}(r,z) = E^{N_1}(r,z) + E^{N_2+1}(r,z) < E^{N_1}(r,z) + E^{N_2}(r,z)$$

which is a contradiction. □

1.4.3 Existence of the mountain pass

In this section, we give some properties of the molecules at infinity implying that the second case of Theorem 1.3 cannot occur. We treat three cases.

1.4.3.1 Case of charged molecules at infinity

In the first case, we assume that the molecules at infinity are charged. This is the simpler case.

Theorem 1.4 (Charged molecules at infinity [45])**.** *Assume that there exist N_1 and N_2 with $N_1 + N_2 = N$ such that*

- $(N_1 - |z|)(N_2 - |z'|) \neq 0$;
- $E^{N_1}(r,z) + E^{N_2}(r',z') = \min\{E^{n_1}(r,z) + E^{n_2}(r',z') \mid n_1 + n_2 = N\}$.

Then the second alternative of Theorem 1.3 does not occur and therefore there exists a critical point at the level c.

The idea of the proof, illustrated in Fig. 1.7, is rather simple: one considers the sequence of paths $\{\gamma_n\}$ furnished by the second alternative of Theorem 1.3 and, using that the molecules attract with each other with a force $(N_1 - |z|)(N_2 - |z'|)/d$, one constructs a new path $\tilde{\gamma}_n$ on which, for n large enough, the highest energy is $< c$, hence getting a contradiction. The new path is constructed on the set of (rotated and translated) tensor products $\psi_1(u\cdot) \wedge \psi_2(u' \cdot -d\vec{v})$ where ψ_1 and ψ_2 are ground states of, respectively, $H^{N_1}(r,z)$ and $H^{N_2}(r',z')$ which are known to exist by Proposition 1.1.

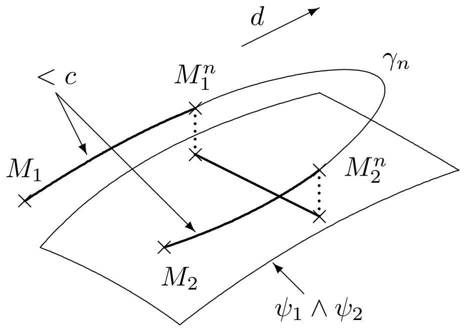


Figure 1.7: Idea of the proof of Theorem 1.4.

1.4.3.2 Case of neutral but polarized molecules at infinity

In the case of neutral molecules at infinity, it is natural to look at the next order in the interaction energy between them, given by the dipole-dipole interaction. The latter depends on the orientation of the molecules and does not have a fixed sign. Let us first recall the definition of the dipole.

Definition 1.1 (Dipole of a molecule in its ground state)**.** *Let* $\Psi \in \bigwedge_1^N L^2(\mathbb{R}^3, \mathbb{C})$ *and* $(R, Z) \in \mathbb{R}^{2M}$*. The corresponding dipole is defined as*

$$P(R, Z, \Psi) := \sum_{i=1}^{N} \int_{(\mathbb{R}^3)^N} x_i |\Psi(x_1, ..., x_N)|^2 dx_1 \cdots dx_N - \sum_{m=1}^{M} Z_m R_m.$$

We say that $H^N(R, Z)$ *has a dipole in its ground state if* $E^N(R, Z) < \Sigma^N(R, Z)$ *and* $P(R, Z, \Psi) \neq 0$ *for all normalized* $\Psi \in ker\{H^N(R, Z) - E^N(R, Z)\} \setminus \{0\}$*.*

The following was proved in [45]:

Theorem 1.5 (Polarized molecules at infinity [45])**.** *Assume that* $|z|$ *and* $|z'|$ *are integers and that*

- $E^{|z|}(r, z) + E^{|z'|}(r', z') < \min\{E^{n_1}(r, z) + E^{n_2}(r', z') \mid n_1 + n_2 = N,\ n_1 \neq |z|\}$;
- *both* $H^{|z|}(r, z)$ *and* $H^{|z'|}(r', z')$ *possess a dipole in their ground state;*
- $E^{|z|}(r, z)$ *or* $E^{|z'|}(r', z')$ *is non degenerated.*

Then the second alternative of Theorem 1.3 does not occur and therefore there exists a critical point at the level c.

The assumption that one of the two ground state energies is non degenerated is a purely technical condition. The proof of Theorem 1.5 is much more involved than the one of Theorem 1.4. An important tool is given by the properties of the dipole-dipole interaction defined as

$$F(x, y) := x \cdot y - 3(x \cdot \vec{v})(y \cdot \vec{v})$$

for $(x, y) \in S^2 \times S^2$. It is proved in [45] that the critical points of F have either a negative energy or a Morse index greater than 1, see Fig. 1.8. Thus either the molecules attract or one can deform the path using the second derivative to decrease the energy.

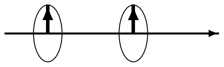
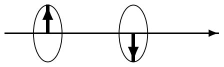
Index	Energy	
4	2	
2	1	
1	-1	
0	-2	

Figure 1.8: Critical points of the dipole-dipole interaction

1.4.3.3 Case when one subsystem is an atom

In [48] it was considered the case for which $m' = 1$, i.e. when one of the two subsystems is a single atom. The following was proved:

Theorem 1.6 (Case of a single atom [48])**.** *Assume that* $m' = 1$*. Then the second alternative of Theorem 1.3 never occurs and therefore there always exists a critical point at the level c.*

This result applies to HCN and many other practical cases. The proof is based on a generalization of the Lieb-Thirring result [53]: one shows that there is always an attractive Van Der Waals interaction between an atom and a molecule, *independently of the orientation of the latter.* The proof is then similar to the case of charged molecules at infinity.

References

[1] *Dalton, a molecular electronic structure program, release 2.0*, 2005. See `http://www.kjemi.uio.no/software/dalton/dalton.html`.

[2] A. Anantharaman and E. Cancès, *On Kohn-Sham models with LDA and GGA exchange-correlation functionals.* `arXiv:0809.5139`, 2008.

[3] T. Ando, *Properties of fermion density matrices*, Rev. Modern Phys., 35 (1963), pp. 690–702.

[4] V. Bach, E. H. Lieb, M. Loss, and J. P. Solovej, *There are no unfilled shells in unrestricted Hartree-Fock theory*, Phys. Rev. Lett., 72 (1994), pp. 2981–2983.

[5] V. Bach, E. H. Lieb, and J. P. Solovej, *Generalized Hartree-Fock theory and the Hubbard model*, J. Statist. Phys., 76 (1994), pp. 3–89.

[6] J. F. Bonnans, J. C. Gilbert, C. Lemaréchal, and C. A. Sagastizábal, *Numerical optimization*, Universitext, Springer-Verlag, Berlin, second ed., 2006. Theoretical and practical aspects.

[7] J. M. Borwein and D. Preiss, *A smooth variational principle with applications to subdifferentiability and to differentiability of convex functions*, Trans. Amer. Math. Soc., 303 (1987), pp. 517–527.

[8] É. Cancès, *Simulation moléculaire et effets d'environnement. Une perspective mathématique et numérique*, PhD thesis, École Nationale des Ponts et Chausées, 1998.

[9] É. Cancès, M. Defranceschi, W. Kutzelnigg, C. Le Bris, and Y. Maday, *Computational quantum chemistry: a primer*, in Handbook of numerical analysis, Vol. X, Handb. Numer. Anal., X, North-Holland, Amsterdam, 2003, pp. 3–270.

[10] É. Cancès, H. Galicher, and M. Lewin, *Computing electronic structures: a new multiconfiguration approach for excited states*, J. Comput. Phys., 212 (2006), pp. 73–98.

[11] É. Cancès and C. Le Bris, *Can we outperform the DIIS approach for electronic structure calculations?*, Int. J. Quantum Chem., 79 (2000), pp. 82–90.

[12] ———, *On the convergence of SCF algorithms for the Hartree-Fock equations*, M2AN Math. Model. Numer. Anal., 34 (2000), pp. 749–774.

[13] É. Cancès, M. Lewin, and G. Stoltz, *The electronic ground-state energy problem: a new reduced density matrix approach.*, J. Chem. Phys., 125 (2006), p. 64101.

[14] S.-L. Chiu, J. McDouall, and I. Hiller, *Prediction of whole reaction paths for large molecular systems*, J. Chem. Soc. Faraday Trans., 90 (1994), pp. 1575–1579.

[15] Y. S. Choi and P. J. McKenna, *A mountain pass method for the numerical solution of semilinear elliptic problems*, Nonlinear Anal., 20 (1993), pp. 417–437.

[16] Y. S. Choi, P. J. McKenna, and M. Romano, *A mountain pass method for the numerical solution of semilinear wave equations*, Numer. Math., 64 (1993), pp. 487–509.

[17] A. Coleman, *Structure of fermion density matrices*, Rev. Modern Phys., 35 (1963), pp. 668–689.

[18] A. Coleman and V. Yukalov, *Reduced Density Matrices: Coulson's Challenge*, Springer Verlag, 2000.

[19] A. J. Coleman, *Kummer variety, geometry of N-representability, and phase transitions*, Phys. Rev. A, 66 (2002), pp. 022503–+.

[20] C. A. Coulson, *Present State of Molecular Structure Calculations*, Reviews of Modern Physics, 32 (1960), pp. 170–177.

[21] R. Czerminski and R. Elber, *Self-avoiding walk between two fixed points as a tool to calculate reaction paths in large molecular systems*, Int. J. Quantum Chemistry, 24 (1990), pp. 167–185.

[22] J. Dolbeault, P. Felmer, and M. Lewin, *Stability of the Hartree-Fock model with temperature*, Math. Models Methods Appl. Sci., 19 (2009), pp. 347–367.

[23] J. Dolbeault, P. Felmer, M. Loss, and E. Paturel, *Lieb-Thirring type inequalities and Gagliardo-Nirenberg inequalities for systems*, J. Funct. Anal., 238 (2006), pp. 193–220.

[24] R. Dreizler and E. Gross, *Density functional theory*, Springer Verlag, 1990.

[25] W. E, W. Ren, and E. vanden-Eijnden, *String method for the study of rare events*, Phys. Rev. B, 66 (2002), pp. 052301–+.

[26] I. Ekeland, *On the variational principle*, J. Math. Anal. Appl., 47 (1974), pp. 324–353.

[27] R. Elber and M. Karplus, *A method for determining reaction paths in large molecules: application to myoglobin*, Chem. Phys. Letters, 139 (1987), pp. 375–380.

[28] G. Fang and N. Ghoussoub, *Second-order information on Palais-Smale sequences in the mountain pass theorem*, Manuscripta Math., 75 (1992), pp. 81–95.

[29] G. Friesecke, *The multiconfiguration equations for atoms and molecules: charge quantization and existence of solutions*, Arch. Ration. Mech. Anal., 169 (2003), pp. 35–71.

[30] ———, *On the infinitude of non-zero eigenvalues of the single-electron density matrix for atoms and molecules*, R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci., 459 (2003), pp. 47–52.

[31] M. Fukuda, B. J. Braams, M. L. Overton, J. K. Percus, M. Yamashita, Z. Zhao, M. Fukuda, B. J. Braams, M. Nakata, M. L. Overton, K. Percus, M. Yamashita, and Z. Zhao, *Large-scale semidefinite programs in electronic structure calculation*, Math. Program., 109 (2007), pp. 553–580.

[32] N. Ghoussoub, *Duality and perturbation methods in critical point theory*, vol. 107 of Cambridge Tracts in Mathematics, Cambridge University Press, Cambridge, 1993. With appendices by David Robinson.

[33] J. T. Golab, D. L. Yeager, and P. Jørgensen, *Multiple stationary point representations in mc scf calculations*, Chemical Physics, 93 (1985), pp. 83 – 100.

[34] G. Henkelman, G. Jóhannesson, and H. Jónsson, *Methods for Finding Saddle Points and Minimum Energy Paths*, Ed. S. D. Schwartz (Kluwer Academic Publishers), 2000.

[35] W. Hunziker, *On the spectra of Schrödinger multiparticle Hamiltonians*, Helv. Phys. Acta, 39 (1966), pp. 451–462.

[36] P. Jørgensen, J. Olsen, and D. L. Yeager, *Generalizations of Newton-Raphson and multiplicity independent Newton-Raphson approaches in multiconfigurational Hartree-Fock theory*, J. Chem. Phys., 75 (1981), pp. 5802–5815.

[37] P. Jørgensen, P. Swanstrøm, and D. L. Yeager, *Guaranteed convergence in ground state multiconfigurational self-consistent field calculations*, J. Chem. Phys., 78 (1983), pp. 347–356.

[38] T. Kato, *Fundamental properties of hamiltonian operators of Schrödinger type*, Trans. Amer. Math. Soc., 70 (1951), pp. 195–221.

[39] W. Kohn and L. J. Sham, *Self-consistent equations including exchange and correlation effects*, Phys. Rev. (2), 140 (1965), pp. A1133–A1138.

[40] K. N. Kudin, G. E. Scuseria, and É. Cancès, *A black-box self-consistent field convergence algorithm: One step closer*, J. Chem. Phys., 116 (2002), p. 8255.

[41] H. Kummer, *n-Representability Problem for Reduced Density Matrices*, J. Math. Phys., 8 (1967), pp. 2063–2081.

[42] C. Le Bris, *Quelques problèmes mathématiques en chimie quantique moléculaire*, PhD thesis, École Polytechnique, 1993.

[43] ———, *A general approach for multiconfiguration methods in quantum molecular chemistry*, Ann. Inst. H. Poincaré Anal. Non Linéaire, 11 (1994), pp. 441–484.

[44] M. Lewin, *The multiconfiguration methods in quantum chemistry: Palais-Smale condition and existence of minimizers*, C. R. Math. Acad. Sci. Paris, 334 (2002), pp. 299–304.

[45] ———, *A mountain pass for reacting molecules*, Ann. Henri Poincaré, 5 (2004), pp. 477–521.

[46] ———, *Solutions of the multiconfiguration equations in quantum chemistry*, Arch. Ration. Mech. Anal., 171 (2004), pp. 83–114.

[47] ———, *Some nonlinear models in Quantum Mechanics*, PhD thesis, University of Paris-Dauphine, June 2004.

[48] ———, *Solution of a mountain pass problem for the isomerization of a molecule with one free atom*, Ann. Henri Poincaré, 7 (2006), pp. 365–379.

[49] ———, *On the computation of excited states with MCSCF methods*, J. Math. Chem., 44 (2008), pp. 967–980. Conference “Mathematical Methods for Ab Initio Quantum Chemistry”, Nice (FRANCE), Nov. 2005.

[50] E. H. Lieb, *Variational principle for many-fermion systems*, Phys. Rev. Lett., 46 (1981), pp. 457–459.

[51] ———, *Bound on the maximum negative ionization of atoms and molecules*, Phys. Rev. A, 29 (1984), pp. 3018–3028.

[52] E. H. Lieb and B. Simon, *The Hartree-Fock theory for Coulomb systems*, Commun. Math. Phys., 53 (1977), pp. 185–194.

[53] E. H. Lieb and W. E. Thirring, *Universal nature of Van Der Waals forces for Coulomb systems*, Phys. Rev. A, 34 (1986), pp. 40–46.

[54] P.-L. Lions, *The concentration-compactness principle in the calculus of variations. The locally compact case, Part I*, Ann. Inst. Henri Poincaré, 1 (1984), pp. 109–149.

[55] ———, *The concentration-compactness principle in the calculus of variations. The locally compact case, Part II*, Ann. Inst. H. Poincaré Anal. Non Linéaire, 1 (1984), pp. 223–283.

[56] ———, *Solutions of Hartree-Fock equations for Coulomb systems*, Commun. Math. Phys., 109 (1987), pp. 33–97.

[57] Y.-K. Liu, M. Christandl, and F. Verstraete, *Quantum Computational Complexity of the N-Representability Problem: QMA Complete*, Phys. Rev. Lett., 98 (2007), pp. 110503–+.

[58] P.-O. Löwdin, *Quantum theory of many-particle systems. I. Physical interpretations by means of density matrices, natural spin-orbitals, and convergence problems in the method of configurational interaction*, Phys. Rev. (2), 97 (1955), pp. 1474–1489.

[59] J. E. Mayer, *Electron Correlation*, Physical Review, 100 (1955), pp. 1579–1586.

[60] D. A. Mazziotti, *Variational minimization of atomic and molecular ground-state energies via the two-particle reduced density matrix*, Phys. Rev. A, 65 (2002), pp. 062511–+.

[61] ———, *First-order semidefinite programming for the direct determination of two-electron reduced density matrices with application to many-electron atoms and molecules*, J. Chem. Phys., 121 (2004), pp. 10957–10966.

[62] ———, *Realization of Quantum Chemistry without Wave Functions through First-Order Semidefinite Programming*, Phys. Rev. Lett., 93 (2004), pp. 213001–+.

[63] ———, *Variational two-electron reduced density matrix theory for many-electron atoms and molecules: Implementation of the spin- and symmetry-adapted T_2 condition through first-order semidefinite programming*, Phys. Rev. A, 72 (2005), pp. 032510–+.

[64] M. McCourt and J. W. McIver Jr, *On the scf calculation of excited states: Singlet states in the two-electron problem*, J. Comput. Chem., 8 (1987), pp. 454–458.

[65] M. Nakata, B. J. Braams, K. Fujisawa, M. Fukuda, J. K. Percus, M. Yamashita, and Z. Zhao, *Variational calculation of second-order reduced density matrices by strong N-representability conditions and an accurate semidefinite programming solver*, J. Chem. Phys., 128 (2008), pp. 164113–+.

[66] J. Olsen, P. Jørgensen, and D. L. Yeager, *Multiconfigurational Hartree-Fock studies of avoided curve crossing using the Newton-Raphson technique*, J. Chem. Phys., 76 (1982), pp. 527–542.

[67] W. Quapp and D. Heidrich, *Analysis of the concept of minimum energy path on the potential energy surface of chemically reacting systems*, Theoret. Chim. Acta, 66 (1984), pp. 245–260.

[68] P. H. Rabinowitz, *Minimax methods in critical point theory with applications to differential equations*, vol. 65 of CBMS Regional Conference Series in Mathematics, Published for the Conference Board of the Mathematical Sciences, Washington, DC, 1986.

[69] M. Reed and B. Simon, *Methods of modern mathematical physics. IV. Analysis of operators*, Academic Press, New York, 1978.

[70] B. O. Roos, *The Complete Active Space Self-Consistent Field method and its applications in electronic structure calculation*, Ab Initio Methods in Quantum Chemistry - II in Adv. Chem. Phys., 69 (1987), pp. 399–446.

[71] M. B. Ruskai, *Absence of discrete spectrum in highly negative ions: II. Extension to fermions*, Commun. Math. Phys., 85 (1982), pp. 325–327.

[72] H. B. Schlegel, *Ab Initio Methods in Quantum Chemistry*, vol. 1, Lawley Ed, 1987, pp. 249–286. and references cited therein.

[73] R. Shepard, *The multiconfiguration self-consistent field method*, Ab Initio Methods in Quantum Chemistry - II in Adv. Chem. Phys., 69 (1987), pp. 63–200.

[74] I. Sigal, *Geometric methods in the quantum many-body problem. Non existence of very negative ions*, Commun. Math. Phys., 85 (1982), pp. 309–324.

[75] I. M. Sigal, *How many electrons can a nucleus bind?*, Annals of Physics, 157 (1984), pp. 307–320.

[76] B. Simon, *Geometric methods in multiparticle quantum systems*, Commun. Math. Phys., 55 (1977), pp. 259–274.

[77] J. P. Solovej, *Proof of the ionization conjecture in a reduced Hartree-Fock model.*, Invent. Math., 104 (1991), pp. 291–311.

[78] ———, *The ionization conjecture in Hartree-Fock theory*, Ann. of Math. (2), 158 (2003), pp. 509–576.

[79] G. Stoltz, *Quelques méthodes mathématiques pour la simulation moléculaire et multiéchelle*, PhD thesis, École Nationale des Ponts et Chaussées, Juin 2007.

[80] S. Teufel, *Adiabatic perturbation theory in quantum dynamics*, vol. 1821 of Lecture Notes in Mathematics, Springer-Verlag, Berlin, 2003.

[81] H. Umeda, M. Sugawara, Y. Fujimura, and S. Koseki, *Multiphoton dissociation dynamics of hydrogen cyanide in nonstationary laser fields: important role of dipole moment function*, Chemical Physics Letters, 229 (1994), pp. 233–238.

[82] C. Van Winter, *Theory of finite systems of particles. I. The Green function*, Mat.-Fys. Skr. Danske Vid. Selsk., 2 (1964).

[83] S. Vugalter and G. M. Zhislin, *Finiteness of a discrete spectrum of many-particle hamiltonians in symmetry spaces (coordinate and momentum representations)*, Teoret. Mat. Fiz., 32 (1977), pp. 70–87.

[84] H.-J. Werner, *Matrix-formulated direct Multiconfiguration Self-Consistent Field and Multiconfiguration Reference Configuration-Interaction methods*, Ab Initio Methods in Quantum Chemistry - II in Adv. Chem. Phys., 69 (1987), pp. 1–62.

[85] H.-J. Werner and P. J. Knowles, *A second order multiconfiguration SCF procedure with optimum convergence*, J. Chem. Phys., 82 (1985), pp. 5053–5063.

[86] H.-J. Werner and W. Meyer, *A quadratically convergent multiconfiguration-self-consistent field method with simultaneous optimization of orbitals and CI coefficients*, J. Chem. Phys., 73 (1980), pp. 2342–2356.

[87] D. Yafaev, *On the point spectrum in the quantum-mechanical many-body problem*, Math. USSR Izv., 40 (1976), pp. 861–896. English translation.

[88] Z. Zhao, B. J. Braams, M. Fukuda, M. L. Overton, and J. K. Percus, *The reduced density matrix method for electronic structure calculations and the role of three-index representability conditions*, J. Chem. Phys., 120 (2004), pp. 2095–2104.

[89] G. M. Zhislin, *Discussion of the spectrum of Schrödinger operators for systems of many particles. (in Russian)*, Trudy Moskovskogo matematiceskogo obscestva, 9 (1960), pp. 81–120.

[90] ———, *A study of the spectrum of the Schrödinger operator for a system of several particles*, Trudy Moskov. Mat. Obšč., 9 (1960), pp. 81–120.

[91] G. M. Zhislin and A. G. Sigalov, *The spectrum of the energy operator for atoms with fixed nuclei on subspaces corresponding to irreducible representations of the group of permutations*, Izv. Akad. Nauk SSSR Ser. Mat., 29 (1965), pp. 835–860.

2

The Hartree-Fock-Bogoliubov theory of neutron stars and white dwarfs

In this chapter, we study a model for an *attractive* system of (pseudo-)relativistic fermions like the ones which are found, for instance, in neutron stars and white dwarfs. The so-called *Hartree-Fock-Bogoliubov* (HFB) model [28, 2] (which generalizes the Hartree-Fock model introduced in the previous chapter) is a very widely used model for attractive quantum systems [28, 3, 10]. In some approximation, it yields to the well-known Bardeen-Cooper-Schrieffer (BCS) theory which has been successfully employed to describe important physical effects such as superconductivity and superfluidity. The HFB model possesses two variables, the same one body density matrix γ as in HF theory, and the pairing density matrix α which always vanishes for purely repulsive systems. The matrix α describes the physical effect of *Cooper pairing*.

We state the existence and the properties of ground states for a simple HFB model. We also show the blow-up in finite time of the solution to the corresponding time-dependent equation, in the case of a purely gravitational force.

2.1 The Chandrasekhar limit

We consider a system of N neutral fermions (like neutrons) which are only submitted to their own gravitational force. We want to take into account relativistic effects and we will use a *pseudo-relativistic kinetic energy* (although one should in principle use the Dirac operator introduced in Chapter 4), defined as

$$T := \sqrt{-c^2\Delta + m^2c^4} - mc^2$$

where m is the mass of each particle and c is the speed of light. For simplicity, we will in this chapter work in a system of units such that $c = 1$.

The Hamiltonian of the system is *formally* defined as

$$H(N,\kappa) := \sum_{k=1}^{N} T_{x_k} - \kappa \sum_{1\le k<\ell\le N} \frac{1}{|x_k - x_\ell|} \tag{2.1}$$

on the fermionic N-body space

$$\bigwedge_1^N L^2(\mathbb{R}^3,\mathbb{C}^q).$$

The integer q is the number of internal degrees of freedom which is 2 for spin-1/2 particles. The coupling constant $\kappa > 0$ parameterizes the strength of the gravitational interaction among the particles and it is very small in practice (if the particles have the mass of a proton, numerically $\kappa \sim 10^{-38}$). The above Hamiltonian can be viewed as a (very simplified) model problem for self-gravitating relativistic Fermi systems which are found, for example, in *neutron stars* and *white dwarfs* [22, 23].

The Hamiltonian (2.1) is not always easy to realize as a self-adjoint operator in particular when κ is too big. However the quadratic form associated with $H(N,\kappa)$ is always well-defined on the Sobolev space $\bigwedge_1^N H^1(\mathbb{R}^3,\mathbb{C}^q)$ by Hardy-Kato's inequality [17, 16]

$$\frac{1}{|x|} \leq \frac{\pi}{2}\sqrt{-\Delta}. \tag{2.2}$$

As we will essentially be interested in the ground state energy, we simply define

$$E(N,\kappa) := \inf\left\{\langle H(N,\kappa)\Psi,\Psi\rangle,\ \Psi \in \bigwedge_1^N H^1(\mathbb{R}^3,\mathbb{C}^q),\ \|\Psi\|_{L^2} = 1\right\}, \tag{2.3}$$

a well-defined quantity which can in principle be equal to $-\infty$.

Lieb, Thirring and Yau have proved in [22, 23] that there exists an integer $N(\kappa)$ such that

$$\begin{cases} E(N,\kappa) > -\infty & \text{when } N \leq N(\kappa), \\ E(N,\kappa) = -\infty & \text{when } N > N(\kappa). \end{cases}$$

The interpretation of this is that very massive stars undergo *variational collapse* in the sense that the energy is not bounded from below, hence there is no minimizer for (2.3), when $N > N(\kappa)$. As stars contain a huge number of particles, it is natural to study the limit $N \to \infty$. Due to the fact that stability holds only when $N \leq N(\kappa)$, this is only possible if at the same time $\kappa \to 0$ (this is because $N(\kappa) \to \infty$ as $\kappa \to 0$). This was done by Lieb, Thirring and Yau in [22, 23], as we will now quickly explain.

The famous *Chandrasekhar functional* was introduced by Chandrasekhar in 1931 [8] (he was a Nobel laureate in physics along with William Alfred Fowler for their work in the theoretical structure and evolution of stars [9]). It reads

$$\mathcal{E}_{\rm Ch}(\rho) := \int_{\mathbb{R}^3} j(\rho(x))\,dx - \frac{\kappa}{2}D(\rho,\rho)$$

where j is the semi-classical kinetic energy

$$j(t) := q(2\pi^2)^{-1}\int_0^{\left(\frac{6\pi^2 t}{q}\right)^{1/3}} p^2\left(\sqrt{m^2+p^2}-m\right)dp. \tag{2.4}$$

and, like in the previous chapter,

$$D(\rho,\rho) := \int_{\mathbb{R}^3}\int_{\mathbb{R}^3} \frac{\rho(x)\rho(y)}{|x-y|}\,dx\,dy.$$

The associated minimization problem is defined by

$$E_{\rm Ch}(N,\kappa) := \inf\left\{\mathcal{E}_{\rm Ch}(\rho) \mid \rho \geq 0,\ \rho \in L^{4/3}(\mathbb{R}^3),\ \int_{\mathbb{R}^3}\rho = N\right\}.$$

It is known [22, 23] that $E_{\rm Ch}(N,\kappa) > -\infty$ if and only if $N \le N_{\rm Ch}(\kappa) = (\tau_c/\kappa)^{3/2}q^{-1/2}$ where

$$\tau_c := 3/2(6\pi^2)^{1/3}\inf\left\{\frac{\int_{\mathbb{R}^3}\rho^{4/3}}{D(\rho,\rho)} \mid \rho \ge 0,\ \rho \in L^{4/3}(\mathbb{R}^3),\ \int_{\mathbb{R}^3}\rho = 1\right\} \simeq 2.677. \tag{2.5}$$

In [23], Lieb and Yau have shown that the largest possible particle number behaves as

$$N(\kappa)\kappa^{3/2} \to \tau_c^{3/2}q^{-1/2} \qquad \text{as } \kappa \to 0 \tag{2.6}$$

and that for any fixed $0 \le \tau < \tau_c q^{-1/3}$, the quantum system is well-approximated by the semiclassical Chandrasekhar energy:

$$\lim_{\substack{N\to\infty\\ \kappa N^{2/3}\to\tau}} \frac{E(N,\kappa)}{E_{\rm Ch}(N,\kappa)} = 1.$$

2.2 The Hartree-Fock-Bogoliubov approximation

When N stays finite, calculating an approximation to the Schrödinger problem (2.3) which is more precise than the Chandrasekhar simple model is of great interest. Like for the atomic case studied in the previous chapter, we will now introduce a nonlinear model.

The Hartree-Fock approximation of the previous Hamiltonian (2.1) reads

$$\mathcal{E}_{\rm HF}(\gamma) = {\rm Tr}(T\gamma) - \frac{\kappa}{2}\iint_{\mathbb{R}^3\times\mathbb{R}^3}\frac{\rho_\gamma(x)\rho_\gamma(y)}{|x-y|}dx\,dy + \frac{\kappa}{2}\iint_{\mathbb{R}^3\times\mathbb{R}^3}\frac{|\gamma(x,y)|^2}{|x-y|}\,dx\,dy$$

where $0 \le \gamma \le 1$ is the one-body density matrix of the fermions, with ${\rm Tr}(\gamma) = N$. The Hartree-Fock-Bogoliubov (HFB) theory furnishes a more precise model in which one has an additional variable α called the *pairing density matrix.* This is described in details in [2] and summarized in an appendix below, see Section 2.5. The energy functional is given by

$$\begin{aligned}\mathcal{E}(\gamma,\alpha) := {\rm Tr}\,(T\gamma) &- \frac{\kappa}{2}\iint_{\mathbb{R}^3\times\mathbb{R}^3}\frac{\rho_\gamma(x)\rho_\gamma(y)}{|x-y|}dx\,dy\\ &+ \frac{\kappa}{2}\iint_{\mathbb{R}^3\times\mathbb{R}^3}\frac{|\gamma(x,y)|^2}{|x-y|}dx\,dy - \frac{\kappa}{2}\iint_{\mathbb{R}^3\times\mathbb{R}^3}\frac{|\alpha(x,y)|^2}{|x-y|}dx\,dy.\end{aligned} \tag{2.7}$$

The variables γ and α are two *operators* acting on $L^2(\mathbb{R}^3;\mathbb{C}^q)$. The operator α is only assumed to be Hilbert-Schmidt, i. e., we have ${\rm Tr}(\alpha^*\alpha) < \infty$. Its kernel is a $q \times q$ matrix which is supposed to be antisymmetric in the following sense: $\alpha(x,y)^T = -\alpha(y,x)$, where ^{T} is the usual transposition of matrices. One has to supplement these conditions with the following operator inequality relating γ and α:

$$\begin{pmatrix}0 & 0\\ 0 & 0\end{pmatrix} \le \begin{pmatrix}\gamma & \alpha\\ \alpha^* & 1-\overline{\gamma}\end{pmatrix} \le \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} \quad \text{on } L^2(\mathbb{R}^3;\mathbb{C}^q)\oplus L^2(\mathbb{R}^3;\mathbb{C}^q). \tag{2.8}$$

This inequality guarantees that the pair (γ,α) is associated to a unique quasi-free state in Fock space [2].

In principle, α could vanish for a minimizer in which case one recovers the usual (generalized) Hartree-Fock model. The kernel $\alpha(x,\sigma;y,\sigma')$ of the operator α should be interpreted as the *two-body wavefunction* of *Cooper pairs.* For more general interactions, the nonvanishing of α is very important to describe physical effects such as superfluidity or superconductivity. It is of course not clear whether such effects do really occur in a system interacting only through the gravitational force as in our study, but this should be seen as a first step towards the study of more complicated systems. Indeed, HFB theory is the model of choice in nuclear physics, for instance [28, 3].

To define the gravitational HFB correctly, we introduce the following (real) Banach space of density matrices

$$\mathcal{X} = \big\{(\gamma,\alpha) \in \mathfrak{S}_1 \times \mathfrak{S}_2 \ : \ \gamma^* = \gamma, \ \alpha^T = -\alpha, \ \|(\gamma,\alpha)\|_{\mathcal{X}} < \infty\big\}, \tag{2.9}$$

equipped with the norm

$$\|(\gamma,\alpha)\|_{\mathcal{X}} = \big\|(1-\Delta)^{1/4}\gamma(1-\Delta)^{1/4}\big\|_{\mathfrak{S}_1} + \big\|(1-\Delta)^{1/4}\alpha\big\|_{\mathfrak{S}_2}. \tag{2.10}$$

We remind the reader that $\mathfrak{S}_1$ and $\mathfrak{S}_2$ denote the space of trace-class and Hilbert-Schmidt operators on $L^2(\mathbb{R}^3;\mathbb{C}^q)$, respectively [29, 27]. Furthermore, we define the following subsets of density matrices in $\mathcal{X}$:

$$\mathcal{K} = \left\{(\gamma,\alpha) \in \mathcal{X} \ : \ \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \le \begin{pmatrix} \gamma & \alpha \\ \alpha^* & 1-\overline{\gamma} \end{pmatrix} \le \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right\}, \tag{2.11}$$

$$\mathcal{K}_\lambda = \{(\gamma,\alpha) \in \mathcal{K} \ : \ \mathrm{Tr}(\gamma) = \lambda\}. \tag{2.12}$$

It can be seen [19] that $\mathcal{E}$ is well-defined on $\mathcal{K}$. The corresponding HFB minimization problem then reads

$$\boxed{I(\lambda) := \inf\big\{\mathcal{E}(\gamma,\alpha) \mid (\gamma,\alpha) \in \mathcal{K}_\lambda\big\}}$$

which could of course be equal to $-\infty$. We now define the largest possible mass $m\lambda^{\mathrm{HFB}}(\kappa)$ of a star in the HFB approximation.

Proposition 2.1 (Boundedness from below and the Chandrasekhar Limit [19])**.** *Let* $m \ge 0$ *and* $0 \le \kappa < 4/\pi$ *be given. Then there exists a unique number* $\lambda^{\mathrm{HFB}}(\kappa) > 0$*, which is independent of* m*, such that the following holds.*

(i) For $0 \le \lambda \le \lambda^{\mathrm{HFB}}(\kappa)$*, we have* $I(\lambda) > -\infty$*.*

(ii) For $\lambda > \lambda^{\mathrm{HFB}}(\kappa)$*, we have* $I(\lambda) = -\infty$*.*

Furthermore, the function $\lambda^{\mathrm{HFB}}(\kappa)$ *is nonincreasing and continuous with respect to* κ*. It satisfies the asymptotic estimate*

$$\lambda^{\mathrm{HFB}}(\kappa) \sim q^{-1/2}\left(\frac{\tau_c}{\kappa}\right)^{3/2} \quad \text{as} \quad \kappa \to 0, \tag{2.13}$$

where τ_c *is the universal constant defined before in* (2.5)*. For any fixed* $\tau < \tau_c\, q^{-1/2}$*, we have*

$$\lim_{\substack{\lambda\to\infty \\ \kappa\lambda^{2/3}\to\tau}} \frac{I(\lambda)}{E_{\mathrm{Ch}}(\lambda,\kappa)} = 1.$$

Hence we see that the semi-classical limit of the HFB model is the same as the original model based on the Schrödinger operator (2.3).

Remark 2.1. *In [19], it is proved that* $I(\lambda) = -\infty$ *for all* $\lambda > 0$ *and all* $m \ge 0$ *when* $\kappa \ge 4/\pi$*. Saying differently, we have* $\lambda^{\mathrm{HFB}}(\kappa) = 0$ *when* $\kappa \ge 4/\pi$*.*

2.3 Existence and properties of ground states

2.3.1 Existence of HFB minimizers

We have seen that for $\lambda \le \lambda^{\mathrm{HFB}}(\kappa)$ the HFB energy is bounded from below, hence a very natural question is the existence of a minimizer for (2.2). The most challenging main feature of the HFB variational problem in gravitational physics is its lack of weak lower semicontinuity (wlsc) due to the attractive interaction among particles [19]. As a consequence of the absence of wlsc, the existence proof for minimizers is much more involved than for the well-studied Hartree-Fock (HF) models arising in atomic physics, where wlsc plays an essential role; see [21, 26]. Another difficulty in the analysis of HFB models stems from its translational invariance and the fact that the main variables are *two operators* related via a complicated constraint inequality. A further complication (although conceptually less important) is the treatment of the pseudo-differential operator describing the kinetic energy of relativistic fermions. Also, we point out that the pairing term α cannot be handled by obvious means: In contrast to the trace-class operator γ, the operator α is *a priori* only Hilbert-Schmidt. Moreover, loosely speaking, the direct and exchange term are both *"subcritical"* in the sense that they may be controlled by the $L^{12/5}$-norm of $\sqrt{\rho_\gamma} \in H^{1/2}(\mathbb{R}^3)$. By contrast, the pairing energy depending on α is *"critical"* because it can only be controlled by the kinetic energy of γ itself. A particular illustration of this difficulty is that even ruling out the vanishing of a minimizing sequence is quite delicate, as we will explain later.

The existence of minimizers with particle numbers below the critical threshold was proved with E. Lenzmann in [19]. Since the functional $\mathcal{E}$ is translation invariant, we have to take into account the unitary action τ_y on $L^2(\mathbb{R}^3)$ given by the group of translations in $\mathbb{R}^3$: $\tau_y f := f(\cdot - y)$. The precise existence result now reads as follows.

Theorem 2.1 (Existence of Minimizers [19])**.** *Fix the integer $q \geq 1$ describing the internal spin degrees of freedom. Furthermore, suppose that $m > 0$ and $0 < \kappa < 1/\pi$. Then, for all $0 < \lambda < \lambda^{\mathrm{HFB}}(\kappa)$, the following properties hold.*

(*i*) *Every minimizing sequence $\{(\gamma_n, \alpha_n)\}_{n\in\mathbb{N}}$ for $I(\lambda)$ is relatively compact in $\mathcal{X}$ up to translations. That is, there is a sequence $\{y_n\}_{n\in\mathbb{N}} \subset \mathbb{R}^3$ such that, after passing to a suitable subsequence, we have*

$$\tau_{y_n}^*(\gamma_n, \alpha_n)\tau_{y_n} \to (\gamma, \alpha) \text{ strongly in } \mathcal{X} \text{ as } n \to \infty,$$

where $(\gamma, \alpha) \in \mathcal{K}_\lambda$ is a minimizer for $I(\lambda)$. In particular, there exists a minimizer (γ, α) for $I(\lambda)$.

(*ii*) *The following binding inequality holds for all $0 < \lambda' < \lambda$:*

$$I(\lambda) < I(\lambda - \lambda') + I(\lambda'),$$

Remark 2.2. *Note that we impose $\kappa < 1/\pi$ instead of the more "natural" condition $\kappa < 4/\pi$. Fortunately, our restriction on κ is practically of no effect, since as we have already mentioned typically $\kappa = Gm^2 \sim 10^{-38}$. Nevertheless, it is an interesting open mathematically question to extend the existence result up to the optimal threshold $\kappa < 4/\pi$.*

Remark 2.3. *As can be checked from our proof, Theorem 2.1 also holds true when the purely gravitational interaction $-1/|x-y|$ is replaced by a general radial potential $W(|x-y|)$ satisfying the following assumptions*

$$|W(|x|)| \leq \frac{1}{|x|} \quad \text{for all } x \in \mathbb{R}^3, \qquad \text{and} \qquad W(|x|) \leq -\frac{\epsilon}{|x|} \quad \text{for } |x| \geq R_0. \tag{2.14}$$

Of course the largest particle number $\lambda^{\mathrm{HFB}}(\kappa)$ has to be defined accordingly.

The proof follows the celebrated *concentration-compactness* method of Pierre-Louis Lions [24, 25], which of course has to be adapted to the case when the main variable is an operator. It consists in studying in details the behavior of a minimizing sequence which is not precompact in $\mathcal{X}$.

The first step is to prove that there is no *vanishing*. Vanishing means that the minimizing sequence always converges weakly to zero when any translation is applied to it. In this case one can show that $\rho_{\gamma_n} \to 0$ strongly in $L^p(\mathbb{R}^3)$ for $1 < p < 3/2$, and that the exchange and direct terms must converge to zero. Therefore in the case of vanishing one has to study the functional

$$(\gamma, \alpha) \in \mathcal{K} \mapsto \operatorname{Tr}(T\gamma) - \frac{\kappa}{2} \iint_{\mathbb{R}^3 \times \mathbb{R}^3} \frac{|\alpha(x, y)|^2}{|x - y|} dx\, dy.$$

In [19] we have proved that this functional is wlsc and we have computed its minimum using ideas of Frank, Lieb, Seiringer and Siedentop in [11]. Finally, we have shown in [19] that this infimum is above $I(\lambda)$ which implies that vanishing cannot occur.

When the minimizing sequence does not encounter vanishing and is not precompact, one has to face the so-called *dichotomy case* of the concentration-compactness principle. We show that the sequence splits into several pieces receding from each other, where at least two parts are relatively compact and carry strictly positive particle numbers $\lambda^1 > 0$ and $\lambda^2 > 0$, respectively. Furthermore, we conclude that the infima $I(\lambda^i)$ with $i = 1, 2$ must be attained, and we find that the ground state energy decomposes as

$$I(\lambda) = I(\lambda^1) + I(\lambda^2) + I(\lambda - \lambda^1 - \lambda^2) = I(\lambda^1 + \lambda^2) + I(\lambda - \lambda^1 - \lambda^2).$$

The last step is, using that $I(\lambda^1)$ and $I(\lambda^2)$ both have minimizers, to prove that actually the binding inequality $I(\lambda^1 + \lambda^2) < I(\lambda^1) + I(\lambda^2)$ must hold. Physically this is very intuitive as when the two minimizers are put very far away, they should interact through a gravitational interaction force. However, in order to mathematically quantify it is necessary to understand the decay properties at infinity of minimizers, when they exist. If γ and α were finite-rank, one could show that they must decay exponentially. However, when they are infinite rank as is expected in our case (see Theorem 2.4 below), it is quite hard to decide what could be the decay of a minimizer. We proved in [19] (see Theorem 2.2 below) by a variational argument that it must decay at least like $1/R^2$. This decay was enough for our purposes.

2.3.2 Properties of minimizers

2.3.2.1 Nonlinear Equation and Decay Estimate

Let us now describe some properties of HFB minimizers (γ, α) obtained in Theorem 2.1. We denote

$$\Gamma = \begin{pmatrix} \gamma & \alpha \\ \alpha^* & 1 - \overline{\gamma} \end{pmatrix}, \tag{2.15}$$

and we introduce the following *HFB mean-field* operator

$$F_\Gamma := \begin{pmatrix} H_\gamma & -\kappa \frac{\alpha(x,y)}{|x-y|} \\ -\kappa \frac{\alpha^*(x,y)}{|x-y|} & -\overline{H_\gamma} \end{pmatrix} \tag{2.16}$$

acting on $L^2(\mathbb{R}^3; \mathbb{C}^q) \oplus L^2(\mathbb{R}^3; \mathbb{C}^q)$. Here

$$H_\gamma := T - \kappa(\rho_\gamma * \frac{1}{|\cdot|})(x) + \kappa \frac{\gamma(x, y)}{|x - y|} \tag{2.17}$$

is the usual mean-field operator of Hartree-Fock theory, see Chapter 1 and [21, 26, 2]. Moreover, it turns to be convenient to define

$$N = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{2.18}$$

Note that Γ commutes with N if and only if $\alpha = 0$, i. e., if and only if the corresponding quasi-free state in the Fock space also commutes with the number operator $\mathcal{N}$, see, e. g., [2].

We can now state some fundamental properties of the minimizers for the HFB model.

Theorem 2.2 (Properties of Minimizers [19])**.** *Let $q \geq 1$ be given and suppose $m > 0$ and $0 < \kappa < 1/\pi$. Assume that $(\gamma, \alpha) \in \mathcal{K}_\lambda$ is a minimizer for $I(\lambda)$ for some $\lambda > 0$. Then there exists a negative real number $\mu < 0$ such that $\Gamma = \Gamma(\gamma, \alpha)$ solves the following nonlinear equation*

$$\Gamma = \chi_{(-\infty,0)} \left(F_\Gamma - \mu N\right) + D, \tag{2.19}$$

where D is a finite rank operator of the same matrix form as Γ and satisfies $\operatorname{ran}(D) \subset \ker(F_\Gamma - \mu N)$.

Moreover, the following decay estimate for the density function holds: for all $R > 0$ sufficiently large, we have

$$\int_{|x| \geq R} \rho_\gamma(x)\, dx \leq \frac{C}{R^2},$$

where $C > 0$ is some constant independent of R.

The decay estimate for $\rho_\gamma(x)$ (and a slightly refined version) plays an important role in the proof of Theorem 2.1, as was explained before.

2.3.2.2 The spin-$1/2$ case ($q = 2$)

Let us now consider the specific case when the number of internal spin degrees of freedom is $q = 2$, which corresponds to the physically relevant case of spin-$1/2$ fermions such as neutrons. We use the same notation as in [1]:

$$\gamma = \begin{pmatrix} \gamma_\uparrow & \gamma_{\uparrow\downarrow} \\ \gamma_{\downarrow\uparrow} & \gamma_\downarrow \end{pmatrix}, \qquad \alpha = \begin{pmatrix} \alpha_\uparrow & \alpha_{\uparrow\downarrow} \\ \alpha_{\downarrow\uparrow} & \alpha_\downarrow \end{pmatrix}.$$

In this case, it was shown by Bach, Fröhlich and Jonsson in [1], based on a concavity result in Lieb [20], that the following holds:

Theorem 2.3 (Reduction to Simpler Minimization Problem [1])**.** *Let $q = 2$ and $(\gamma, \alpha) \in \mathcal{K}$. Then we have*

$$\mathcal{E}(\gamma, \alpha) \geq \mathcal{E}\left(\tau \otimes \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \sqrt{\tau(1-\tau)} \otimes \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\right) \tag{2.20}$$

where

$$\tau = \frac{\gamma_\uparrow + \gamma_\downarrow + \overline{\gamma_\uparrow} + \overline{\gamma_\downarrow}}{4}.$$

Additionally, there is equality in (2.20) if and only if

$$\gamma = \tau \otimes \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad \text{and} \qquad \alpha = \pm\sqrt{\tau(1-\tau)} \otimes \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Theorem 2.3 is contained in [1]. In particular it means that a minimizer of the HFB minimization problem is always a projection. It also implies the following important equality

$$I(\lambda) = \inf\left\{\mathcal{E}\left(\tau \otimes \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \sqrt{\tau(1-\tau)} \otimes \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\right), \quad 0 \leq \tau \leq 1,\ \tau = \tau^* = \overline{\tau},\ \operatorname{Tr}(\tau) = \lambda/2\right\}. \tag{2.21}$$

This leads to the study of the (no-spin) energy being defined as

$$\tau \mapsto \operatorname{Tr}\left(T\tau\right) - \kappa D(\rho_\tau, \rho_\tau) + \frac{\kappa}{2} \iint_{\mathbb{R}^3 \times \mathbb{R}^3} \frac{|\tau(x,y)|^2}{|x-y|} dx\, dy - \frac{\kappa}{2} \iint_{\mathbb{R}^3 \times \mathbb{R}^3} \frac{\left|\sqrt{\tau(1-\tau)}(x,y)\right|^2}{|x-y|} dx\, dy. \tag{2.22}$$

We note that this energy has a form similar to the Müller functional which was studied in [11]. Indeed, a straightforward adaptation of an argument given in [11] leads to the following statement.

Theorem 2.4 (Infinite Rank of Minimizers if $\alpha \neq 0$, [19])**.** *Assume that (γ, α) is a minimizers for $I(\lambda)$ with $0 < \kappa < 4/\pi$, $\lambda > 0$ and $q = 2$. Then, if $\alpha \neq 0$ holds, the operators γ and α both have infinite rank.*

2.4 The Dynamical Collapse

As it is well-known from nonlinear Schrödinger equations, a detailed understanding of the variational calculus for the time-independent theory turns out to be of great use when addressing dynamical questions; e. g., the stability of solitary waves and blowup analysis.

For the reader's orientation, we recall that the time-dependent HFB equations generated by the functional $\mathcal{E}(\gamma,\alpha)$ can be written in commutator form as

$$\begin{cases} i\dfrac{\partial}{\partial_t}\Gamma = [F_\Gamma, \Gamma], \\ \Gamma(t=0) = \Gamma_0 \in \mathcal{K}. \end{cases} \tag{2.23}$$

Here and as usual, $\Gamma = \Gamma(\gamma,\alpha)$ denotes the admissible 1-pdm defined in (2.15) and F_Γ is the HFB mean-field operator introduced in (2.16). Clearly, the equation (2.23) is a nonlinear evolution equation. With regard to the wellposedness of its initial-value problem, we note that a straightforward adaptation of [7, 6, 4, 13, 15, 18] yields the following result.

Theorem 2.5 (Well-posedness). *Assume that $\kappa \geq 0$. For each initial datum $(\gamma_0,\alpha_0) \in \mathcal{K}$, there exists a unique maximal solution $(\gamma,\alpha) \in C^0([0,t_{\max});\mathcal{X}) \cap C^1([0,t_{\max});\mathcal{X}')$ solving (2.23) such that $(\gamma(0),\alpha(0)) = (\gamma_0,\alpha_0)$. Moreover, we have conservation of energy and expected number of particles, i. e.,*

$$\forall 0 \leq t < t_{\max}, \qquad \mathcal{E}(\gamma(t),\alpha(t)) = \mathcal{E}(\gamma_0,\alpha_0) \quad \textit{and} \quad \mathrm{Tr}(\gamma(t)) = \mathrm{Tr}(\gamma_0).$$

If the initial datum $(\gamma_0,\alpha_0) \in \mathcal{K}$ satisfies the smallness condition $\mathrm{Tr}(\gamma_0) < \lambda^{\mathrm{HFB}}(\kappa)$, then we have $t_{\max} = +\infty$ and $\Gamma(t)$ is bounded in $\mathcal{X}$.

It is straightforward to verify that if Γ_0 is an orthogonal projector on $L^2(\mathbb{R}^3,\mathbb{C}^q) \oplus L^2(\mathbb{R}^3,\mathbb{C}^q)$, then so is $\Gamma(t)$ for all times $t \in [0,t_{\max})$. Similarly, if $\Gamma_0 = (\gamma_0, 0)$ is a Hartree-Fock state, then $\alpha(t) \equiv 0$ for all times.

Next, we consider the behavior of minimizers (when they exist) for the HFB energy $\mathcal{E}(\gamma,\alpha)$ with respect to the evolution equation (2.23). To this end, let us assume that the initial condition (γ_0,α_0) is a minimizer for $I(\lambda)$ as given by Theorem 2.2. With this choice of initial conditions and using equation (2.19) above, an elementary calculation shows that the corresponding solution $\Gamma(t) = \Gamma(\gamma(t),\alpha(t))$ of (2.23) is given by

$$\gamma(t) = \gamma_0 \quad \text{and} \quad \alpha(t) = e^{-2\mu i t}\alpha_0, \tag{2.24}$$

or equivalently in a 2×2 matrix form

$$\Gamma(t) = e^{i\mu N t}\,\Gamma_0\,e^{-i\mu N t} \tag{2.25}$$

with $\mu < 0$ taken from Theorem 2.2 and N defined in (2.18). Hence HFB minimizers give rise to stationary solutions of (2.23), as one naturally expects. Moreover, by adapting a well-known general argument in [5], the relative compactness result of Theorem 2.1 and the conservation laws imply orbital stability of minimizers. The precise statement can be formulated as follows. Let us define the set of minimizers

$$\mathcal{M}_\lambda = \big\{(\gamma,\alpha) \in \mathcal{K}_\lambda : \mathcal{E}(\gamma,\alpha) = I(\lambda)\big\},$$

and introduce the distance function on $\mathcal{K}$

$$\mathrm{dist}_{\mathcal{M}_\lambda}(\gamma_0,\alpha_0) = \inf_{(\gamma,\alpha)\in\mathcal{M}_\lambda} \|(\gamma_0,\alpha_0) - (\gamma,\alpha)\|_{\mathcal{X}}.$$

Theorem 2.6 (Stability of Minimizers under HFB Time Evolution [19]). *Under the assumptions of Theorem 2.1, the set of minimizers $\mathcal{M}_\lambda$ for $I(\lambda)$ with $0 < \lambda < \lambda^{\rm HFB}(\kappa)$ is* orbitally stable *in the following sense: for every $\epsilon > 0$, there is $\delta > 0$ such that if the initial condition satisfies $\operatorname{dist}_{\mathcal{M}_\lambda}(\gamma_0, \alpha_0) < \delta$, then the corresponding solution $\Gamma(t) = (\gamma(t), \alpha(t))$ of (2.23) exists for all times $t \geq 0$ and obeys $\sup_{t\geq 0} \operatorname{dist}_{\mathcal{M}_\lambda}(\gamma(t), \alpha(t)) < \epsilon$.*

Finally, we turn to the study of the finite-time blowup for the time-dependent HFB equation, when the initial condition has a large number of particles (i.e. when the star has a large mass). This models the *dynamical collapse* of stars which are too massive.

Theorem 2.7 (Finite-time blow-up in Generalized HF theory [14]). *Let $0 \leq \gamma_0 \leq 1$ be a spherically symmetric one-body density matrix, i.e. such that $\gamma_0(Rx, Ry) = \gamma_0(x, y)$ for all $R \in SO(3)$. Assume moreover that*

$$\mathrm{Tr}(\gamma_0) + \mathrm{Tr}(\gamma_0 |L|^2) < \infty$$

where $L = x \times (-i\nabla)$ is the angular momentum. If

$$\mathcal{E}(\gamma_0, 0) < -m\mathrm{Tr}(\gamma_0), \tag{2.26}$$

then the maximal solution of the time-dependent equation (2.23) with initial datum $(\gamma_0, 0)$ blows up in finite time: one has $0 < t_{\max} < \infty$ and

$$\lim_{t\to t_{\max}} \mathrm{Tr}\big((-\Delta)^{1/4}\gamma(t)(-\Delta)^{1/4}\big) = \infty.$$

We note that (2.26) requires that $\mathrm{Tr}(\gamma_0)$ is large enough, at least $\mathrm{Tr}(\gamma_0) > \lambda^{\rm HF}(\kappa)$. As the initial state is a Hartree-Fock state ($\alpha_0 = 0$), it stays a Hartree-Fock state for all times.

A result similar to Theorem 2.7 was proved first by Fröhlich and Lenzmann in [13, 12] for simplified models (Choquard functional or neglecting the exchange term in HF theory). Hainzl and Schlein have proved Theorem 2.7 in the particular case for which γ_0 is a N-dimensional orthogonal projector, $\gamma_0^2 = \gamma_0$. Their proof uses as an essential tool that $\gamma(t)$ stays a projector, hence is finite rank for all times. The generalization to any one-body density matrix was tackled in [14]. A very interesting open question is to extend this result to non-vanishing pairing, $\alpha_0 \neq 0$.

2.5 Appendix: Hartree-Fock-Bogoliubov Theory

Here we briefly recap Hartree-Fock-Bogoliubov (HFB) theory as is detailed in [2], for the convenience of the reader and also because we want to relate this to the models of the next chapters. Some of the material of this section is repeated and detailed in a more general setting in Appendix A. A key observation is the fact the the set of admissible pairs (γ, α), which appears in our main problem (2.2), stays in one-to-one correspondence to the set *quasi-free states* ω in Fock space, having a finite particle number, which generalize Slater determinants of the usual Hartree-Fock approximation.

Fock Space and Full Hamiltonian

We begin, rather abstractly, by considering a general microscopic model of a quantum system described in terms of a second-quantized Hamiltonian $\mathbb{H}$ acting on a fermionic Fock space $\mathcal{F}$, which is defined as follows. Let $\mathfrak{H}$ be a one-particle Hilbert space (finite or infinite-dimensional, but always separable). For each integer $N \geq 1$, we define the fermionic N-particle space $\mathfrak{H}^{(N)}$ as the N-fold antisymmetric tensor product

$$\mathfrak{H}^{(N)} = \bigwedge_1^N \mathfrak{H}, \quad \text{for } N = 1, 2, 3, \ldots \tag{2.27}$$

The fermionic *Fock space* $\mathcal{F}$ is defined to be the infinite direct sum of Hilbert spaces

$$\mathcal{F} = \mathfrak{H}^{(0)} \oplus \mathfrak{H}^{(1)} \oplus \mathfrak{H}^{(2)} \oplus \mathfrak{H}^{(3)} \oplus \dots . \tag{2.28}$$

Here, by definition, we set $\mathfrak{H}^{(0)} = \mathbb{C}$ with $\Omega = 1 \oplus 0 \oplus \cdots$ being called the *vacuum.* To any $f \in \mathfrak{H}$ we associate a creation operator $a^\dagger(f)$ and an annihilation operator $a(f)$, both acting on $\mathcal{F}$, as follows. The action of $a^\dagger(f)$ on simple vectors is defined as

$$a^\dagger(f)(f_1 \wedge \cdots \wedge f_N) = f \wedge f_1 \wedge \cdots \wedge f_N. \tag{2.29}$$

By linearity, this definition extends to all of $\mathcal{F}$, and $a(f) := (a^\dagger(f))^*$ is defined as the adjoint of the operator $a^\dagger(f)$ on $\mathcal{F}$. Note that $a^\dagger(f)$ and $a(f)$ are bounded operators, and we have $a(f)\Omega = 0$ and $a(\lambda f) = \overline{\lambda} a(f)$ for any $f \in \mathfrak{H}$ and any $\lambda \in \mathbb{C}$. Furthermore, one verifies that $a^\dagger(f)$ and $a(f)$ obey the *canonical anticommutation relations (CAR):* For all $f, g \in \mathfrak{H}$, we have

$$\begin{gathered} \{a(f), a^\dagger(g)\} = \langle f, g \rangle \mathbb{1}, \\ \{a^\dagger(f), a^\dagger(g)\} = \{a(f), a(g)\} = 0. \end{gathered} \tag{2.30}$$

Here $\{A, B\} = AB + BA$ denotes the anticommutator, $\langle f, g \rangle$ is the inner product on $\mathfrak{H}$ (with the convention that its conjugate linear in f), and $\mathbb{1}$ stands for the identity map on $\mathcal{F}$. We assume that $\{f_i\}$ is an orthonormal basis of $\mathfrak{H}$ and we set $a_i = a(f_i)$ in the following.

Let us now consider an N-body Hamiltonian acting on $\mathfrak{H}^{(N)}$, composed of a one-body and a two-body term:

$$H^N := \sum_{k=1}^N h_k + \frac{1}{2} \sum_{1 \le k, \ell \le N} V_{k\ell}$$

where $h : \mathfrak{H} \to \mathfrak{H}$ and $V : \mathfrak{H} \wedge \mathfrak{H} \to \mathfrak{H} \wedge \mathfrak{H}$. The family of Hamiltonians H^N acting on $\mathfrak{H}^{(N)}$ gives rise to a Hamiltonian $\mathbb{H}$ in Fock space which is simply

$$\mathbb{H} := 0 \oplus \bigoplus_{N \ge 1} H^N.$$

It can be verified that it holds

$$\mathbb{H} = \sum_{i,j} h_{ij} a_i^\dagger a_j + \frac{1}{2} \sum_{k,l,m,n} V_{kl;mn} a_k^\dagger a_l^\dagger a_n a_m \tag{2.31}$$

where the matrix elements of h and V are given by

$$h_{ij} = \langle f_i, h f_j \rangle, \qquad V_{kl;mn} = \frac{1}{2} \langle f_k \wedge f_l, V f_m \wedge f_n \rangle. \tag{2.32}$$

At this point we will treat H^N and $\mathbb{H}$ as formal objects. In practice they of course have to be realized as self-adjoint operators on $\mathfrak{H}^{(N)}$ and $\mathcal{F}$.

Next, we recall that a *state* ω on $\mathcal{F}$ is a $\mathbb{C}$-valued linear map defined on $\mathcal{B}(\mathcal{F})$ (the set of bounded operators on $\mathcal{F}$) such that

$$\omega(\mathbb{1}) = 1 \quad \text{and} \quad \omega(A^* A) \ge 0, \quad \text{for all } A \in \mathcal{B}(\mathcal{F}). \tag{2.33}$$

Note that the latter condition implies that $\overline{\omega(A)} = \omega(A^*)$. In particular, if $\mathcal{D} \in \mathfrak{S}_1(\mathcal{F})$ is a nonnegative trace-class operator with $\mathrm{Tr}_{\mathcal{F}}(\mathcal{D}) = 1$, then setting

$$\omega(A) = \mathrm{Tr}(\mathcal{D} A), \quad \text{for all } A \in \mathcal{B}(\mathcal{F}), \tag{2.34}$$

clearly defines a state in $\mathcal{F}$. States of this form are usually called *normal states.* The simplest normal states are the so-called *pure states* given by $\mathcal{D} = |\Psi\rangle\langle\Psi|$ for some normalized $\Psi \in \mathcal{F}$. If for instance $\Psi = 0 \oplus \cdots \oplus \psi \oplus 0 \oplus \cdots$ with $\psi \in \mathfrak{H}^{(N)}$, then we just get $\omega(\mathbb{H}) = \left\langle H^N \psi, \psi \right\rangle_{\mathfrak{H}^{(N)}}$. Note that we can extend the action of normal states $\omega = \mathrm{Tr}(\mathcal{D}\cdot)$ to operators which are bounded below. Indeed, by assuming without loss of generality that $A \geq 0$ holds, we see that $\omega(A) = \mathrm{Tr}(\mathcal{D}_\omega A)$ is well-defined (but possibly equals $+\infty$). An important example for such an semibounded operator is given by the *number operator*

$$\mathcal{N} = \bigoplus_{N \geq 0} N = \sum_{i \geq 1} a_i^\dagger a_i. \tag{2.35}$$

We say that a state ω has *finite particle number* if

$$\omega(\mathcal{N}) = \sum_{N=0}^{\infty} N \omega(\Pi^{(N)}) < \infty \tag{2.36}$$

where $\Pi^{(N)}$ is the projection onto the subspace $\mathfrak{H}^{(N)} \subset \mathcal{F}$. Such states are of primary physical importance and they lead to the variational problem

$$E(N) = \inf\{\omega(\mathbb{H}) : \omega(\mathcal{N}) = N\}, \tag{2.37}$$

which corresponds to finding the (formal) ground state energy of $\mathbb{H}$ subject to the constraint of a fixed average particle number N.

For any state ω, we may define the *density matrices* γ and α. Following [30], let us denote by $\mathcal{J} : \mathfrak{H} \to \mathfrak{H}^*$ the dual (anti-linear) map defined as $\mathcal{J}(f)(g) = \langle f, g \rangle$. When $\mathfrak{H} = L^2(M, \mathbb{C}^q)$ for some measurable set M in $\mathbb{R}^d$, one can simply identify $\mathcal{J}$ to the conjugation operator. We then view α as a linear operator defined on $\mathcal{J}\mathfrak{H} = \mathfrak{H}^*$ with values in $\mathfrak{H}$:

$$\langle g, \gamma f \rangle = \omega\left(a^\dagger(f) a(g)\right), \qquad \langle g, \alpha \mathcal{J} f \rangle = \omega\left(a(f) a(g)\right). \tag{2.38}$$

the operator $\gamma : \mathfrak{H} \to \mathfrak{H}$ is usually called the *one-body density matrix*, whereas $\alpha : \mathfrak{H} \to \mathfrak{H}^*$ is called the *pairing density matrix*.

From γ and α we can construct an operator Γ acting on $\mathfrak{H} \oplus \mathcal{J}\mathfrak{H}$ dy defining the so-called *generalized one-body density matrix*

$$\Gamma = \begin{pmatrix} \gamma & \alpha \\ \alpha^* & 1 - \mathcal{J}\gamma\mathcal{J}^* \end{pmatrix}. \tag{2.39}$$

It can be shown using the CAR (2.30) that Γ is a self-adjoint bounded operator satisfying $0 \leq \Gamma \leq 1$ and that α is antisymmetric in the sense that $\langle g, \alpha \mathcal{J} f \rangle = -\langle f, \alpha \mathcal{J} g \rangle$.

When $\omega(\mathcal{N}) < \infty$, $\gamma \geq 0$ is trace-class and $\omega(\mathcal{N}) = \mathrm{Tr}(\gamma)$. In this case $\mathrm{Tr}(\alpha^* \alpha) < \infty$, i.e. α must be a Hilbert-Schmidt operator. Indeed α can be identified with a two-body state $\hat{\alpha}$ in $\mathfrak{H} \wedge \mathfrak{H}$ through the property $\langle g, \alpha \mathcal{J} f \rangle = \langle g \wedge f, \hat{\alpha} \rangle_{\mathfrak{H} \wedge \mathfrak{H}}$. We remark that Γ itself does not have finite trace, but we always have $\mathrm{Tr}(\Gamma(1 - \Gamma)) < \infty$.

HFB Approximation and Quasi-Free States

When studying the variational problem (2.37), one often has to introduce suitable approximations. For instance, the Hartree-Fock (HF) method consists of restricting to states given by *N-particle Slater determinants*. That is, states ω are of the form $\omega(A) = \langle \Psi, A\Psi \rangle$, where $\Psi \in \mathfrak{H}^{(N)}$ is a simple vector such that

$$\Psi = a^\dagger(f_1) \cdots a^\dagger(f_N) \Omega, \quad \text{for some } f_i \in \mathfrak{H} \text{ with } \langle f_i, f_j \rangle = \delta_{ij}. \tag{2.40}$$

A further generalization of the HF method is given by the so-called *Hartree-Fock-Bogoliubov theory* (or *generalized Hartree-Fock method* [2]), which plays a very important role for *attractive* two-body potentials; e. g., in nuclear physics, superconductivity, and stellar physics. In HFB theory, the set of admissible ω is significantly enlarged to the set of so-called quasi-free states. By definition, we say that a state ω is *quasi-free* if, for any integer $N \geq 1$,

$$\omega(a_1^\# a_2^\# \cdots a_{2N}^\#) = \sum_\pi (-)^\pi \omega(a_{\pi(1)}^\# a_{\pi(2)}^\#) \cdots \omega(a_{\pi(2N-1)}^\# a_{\pi(2N)}^\#). \tag{2.41}$$

Here $a_1^\#, a_2^\#, \dots, a_{2N}^\#$ are each either $a^\dagger$ or a, and the sum $\sum_\pi$ runs over all permutations π which satisfy $\pi(1) < \pi(3) < \cdots < \pi(2N-1)$ and $\pi(2j-1) < \pi(2j)$ for all $1 \leq j \leq N$. The condition (2.41) means that any expectation value can be calculated in terms of only the one-body matrices γ and α. The formula (2.41) indeed generalizes a property of Slater determinants which can easily be seen to be quasi-free.

Having defined quasi-free states, we introduce the Hartree-Fock-Bogoliubov (HFB) minimization problem as

$$E_{\mathrm{HFB}}(N) = \inf\{\omega(\mathbb{H}) : \omega(\mathcal{N}) = N \text{ and } \omega \text{ is quasi-free}\}. \tag{2.42}$$

Clearly, we have that $E(N) \leq E_{\mathrm{HFB}}(N)$ holds. Moreover, by recalling that $\mathbb{H}$ is of the form (2.31) and by using (2.41) with $N = 2$, we can rewrite HFB minimization problem as follows.

$$E_{\mathrm{HFB}}(N) = \inf\big\{\mathcal{E}_{\mathrm{HFB}}(\Gamma) : \Gamma \text{ is an admissible one-body density matrix, } \mathrm{Tr}(\gamma) = N\big\}. \tag{2.43}$$

where

$$\mathcal{E}_{\mathrm{HFB}}(\Gamma) = \mathrm{Tr}(h\gamma) + \frac{1}{2}\sum_{k,l,m,n} V_{klmn}(\gamma_{mk}\gamma_{nl} - \gamma_{ml}\gamma_{nk} + \alpha_{lk}^*\alpha_{mn}), \tag{2.44}$$

with matrix elements $\gamma_{kl} = \langle f_i, \gamma f_j\rangle$ and $\alpha_{ij} = \langle f_i, \alpha f_j\rangle$. The last three terms in (2.44) are called *direct, exchange,* and *pairing energy* respectively.

By definition, we say that Γ is an *admissible one-body density matrix* if it arises from a quasi-free state having a finite particle number. It was shown in [2] that essentially all density matrices are admissible:

Theorem 2.8 (Hartree-Fock-Bogoliubov states are characterized by the density matrix [2]). *Let Γ be a self-adjoint operator acting on $\mathfrak{H} \oplus \mathfrak{H}^*$ of the form* (2.39) *such that $0 \leq \Gamma \leq 1$, $\mathrm{Tr}(\gamma) < \infty$ and $\langle g, \alpha\mathcal{J}f\rangle = -\langle f, \alpha\mathcal{J}g\rangle$ for all $f, g \in \mathfrak{H}$. Then there exists a unique quasi-free state ω on $\mathcal{F}$ with finite particle number, whose generalized one-body particle density matrix is Γ.*

HFB Energy for Self-Gravitating Gases

For fermionic systems interacting through the gravitational interaction, one takes

$$\mathfrak{H} = L^2(\mathbb{R}^3, \mathbb{C}^q), \qquad h = T = \sqrt{-\Delta + m^2} - m, \qquad V(x,y) = -\frac{\kappa}{|x-y|}.$$

This leads to (2.7).

References

[1] V. Bach, J. Fröhlich, and L. Jonsson, *Bogolubov-Hartree-Fock mean field theory for neutron stars and other systems with attractive interactions.* Preprint `mp-arc:08-243`, 2008.

[2] V. Bach, E. H. Lieb, and J. P. Solovej, *Generalized Hartree-Fock theory and the Hubbard model*, J. Statist. Phys., 76 (1994), pp. 3–89.

[3] M. Bender, P.-H. Heenen, and P.-G. Reinhard, *Self-consistent mean-field models for nuclear structure*, Reviews of Modern Physics, 75 (2003), pp. 121–180.

[4] A. Bove, G. Da Prato, and G. Fano, *On the Hartree-Fock time-dependent problem*, Commun. Math. Phys., 49 (1976), pp. 25–33.

[5] T. Cazenave and P.-L. Lions, *Orbital stability of standing waves for some nonlinear Schrödinger equations*, Commun. Math. Phys., 85 (1982), pp. 549–561.

[6] J. M. Chadam, *The time-dependent Hartree-Fock equations with Coulomb two-body interaction*, Commun. Math. Phys., 46 (1976), pp. 99–104.

[7] J. M. Chadam and R. T. Glassey, *Global existence of solutions to the Cauchy problem for time-dependent Hartree equations*, J. Math. Phys., 16 (1975), pp. 1122–1130.

[8] S. Chandrasekhar, *The maximum mass of ideal white dwarfs*, Astrophys. J., 74 (1931), pp. 81–82.

[9] ———, *On stars, their evolution and their stability*, 1983. Nobel lecture.

[10] D. J. Dean and M. Hjorth-Jensen, *Pairing in nuclear systems: from neutron stars to finite nuclei*, Rev. Mod. Phys., 75 (2003), pp. 607–656.

[11] R. L. Frank, E. H. Lieb, R. Seiringer, and H. Siedentop, *Müller's exchange-correlation energy in density-matrix-functional theory*, Phys. Rev. A, 76 (2007), p. 052517.

[12] J. Fröhlich and E. Lenzmann, *Blowup for nonlinear wave equations describing boson stars*, Comm. Pure Appl. Math., 60 (2007), pp. 1691–1705.

[13] ———, *Dynamical collapse of white dwarfs in Hartree- and Hartree-Fock theory*, Commun. Math. Phys., 274 (2007), pp. 737–750.

[14] C. Hainzl, E. Lenzmann, M. Lewin, and B. Schlein. in preparation, 2009.

[15] C. Hainzl, M. Lewin, and C. Sparber, *Existence of global-in-time solutions to a generalized Dirac-Fock type evolution equation*, Lett. Math. Phys., 72 (2005), pp. 99–113.

[16] I. W. Herbst, *Spectral theory of the operator* $(p^2 + m^2)^{1/2} - Ze^2/r$, Commun. Math. Phys., 53 (1977), pp. 285–294.

[17] T. Kato, *Perturbation theory for linear operators*, Springer, second ed., 1995.

[18] E. Lenzmann, *Well-posedness for semi-relativistic Hartree equations of critical type*, Math. Phys. Anal. Geom., 10 (2007), pp. 43–64.

[19] E. Lenzmann and M. Lewin, *Minimizers for the Hartree-Fock-Bogoliubov theory of neutron stars and white dwarfs*. `arXiv:0809.2560`, 2008.

[20] E. H. Lieb, *Convex trace functions and the Wigner-Yanase-Dyson conjecture*, Advances in Math., 11 (1973), pp. 267–288.

[21] E. H. Lieb and B. Simon, *The Hartree-Fock theory for Coulomb systems*, Commun. Math. Phys., 53 (1977), pp. 185–194.

[22] E. H. Lieb and W. E. Thirring, *Gravitational collapse in quantum mechanics with relativistic kinetic energy*, Ann. Physics, 155 (1984), pp. 494–512.

[23] E. H. Lieb and H.-T. Yau, *The Chandrasekhar theory of stellar collapse as the limit of quantum mechanics*, Commun. Math. Phys., 112 (1987), pp. 147–174.

[24] P.-L. Lions, *The concentration-compactness principle in the calculus of variations. The locally compact case, Part I*, Ann. Inst. Henri Poincaré, 1 (1984), pp. 109–149.

[25] ———, *The concentration-compactness principle in the calculus of variations. The locally compact case, Part II*, Ann. Inst. H. Poincaré Anal. Non Linéaire, 1 (1984), pp. 223–283.

[26] ———, *Solutions of Hartree-Fock equations for Coulomb systems*, Commun. Math. Phys., 109 (1987), pp. 33–97.

[27] M. Reed and B. Simon, *Methods of modern mathematical physics. IV. Analysis of operators*, Academic Press, New York, 1978.

[28] P. Ring and P. Schuck, *The nuclear many-body problem*, vol. Texts and Monographs in Physics, Springer Verlag, New York, 1980.

[29] B. Simon, *Trace ideals and their applications*, vol. 35 of London Mathematical Society Lecture Note Series, Cambridge University Press, Cambridge, 1979.

[30] J. P. Solovej, *Many body quantum mechanics*. Lectures notes at LMU Münich, 2007.

3

The thermodynamic limit of Coulomb quantum systems

This chapter is almost identical to a review [15] of the two papers [16, 17], written in collaboration with C. Hainzl and J.P. Solovej. It deals with the study of quantum systems interacting through Coulomb forces when the number of particles goes to infinity (thermodynamic limit).

3.1 Introduction

Two difficulties arise when describing particles interacting through the Coulomb potential (as the ones in ordinary matter). Both have to do with the physical problem of *stability* of quantum systems.

The first is due to the singularity of $1/|x|$ at 0: it is necessary to explain why a particle will not rush to a particle of the opposite charge. One of the first major triumphs of the theory of quantum mechanics is the explanation it gives of the stability of the hydrogen atom (and the complete description of its spectrum) and of other microscopic quantum Coulomb systems, via the uncertainty principle. One often refers to this kind of stability as *stability of the first kind* [20, 21]. If we denote by $E(N)$ the ground state energy of the system under consideration, for N particles, stability of the first kind can be written

$$E(N) > -\infty. \tag{3.1}$$

In proving (3.1) for Coulomb systems, a major role is played by the uncertainty principle which for nonrelativistic systems is mathematically expressed by the critical Sobolev embedding $H^1(\mathbb{R}^3) \hookrightarrow L^6(\mathbb{R}^3)$. The latter allows to prove Kato's inequality

$$\forall \epsilon > 0, \qquad \frac{1}{|x|} \leq \epsilon(-\Delta) + \frac{1}{\epsilon},$$

which means that the Coulomb potential is controlled by an arbitrary small amount of kinetic energy.

The second issue concerns the slow decay of $1/|x|$ at infinity and this has to do with the macroscopic behavior of quantum Coulomb systems, as we will see in details in this chapter. It is indeed necessary to explain how a very large number of electrons and nuclei can stay bounded together to form macroscopic systems, although each particle interacts with a lot of other charged particles due to the long tail of the Coulomb interaction potential. Whereas the stability of atoms was an early triumph of quantum mechanics it, surprisingly, took nearly forty years before the question of stability of everyday macroscopic objects was even raised (see Fisher and Ruelle [11]). A rigorous answer to the question came shortly thereafter in what came to be known as the Theorem on Stability of Matter proved first by Dyson and Lenard [8].

The main question which we want to address is how the lowest possible energy $E(N)$ appearing in (3.1) depends on the (macroscopic) number N of particles in the object. More precisely, one is interested in proving a behavior of the form

$$E(N) \sim_{N\to\infty} \bar{e}N. \tag{3.2}$$

This behavior as the number of particles grows is mandatory to explain why matter does not collapse or explode in the thermodynamic limit. Assume that (3.2) does not hold and that for instance $E(N) \sim_{N\to\infty} cN^p$ with $p \neq 1$. Then $|E(2N) - 2E(N)|$ becomes very large as $N \gg 1$. Depending on p and the sign of the constant c, a very large amount of energy will be either released when two identical systems are put together, or necessary to assemble them. The constant $\bar{e}$ in (3.2) is the energy per particle.

Stability of Matter is itself a necessary first step towards a proof of (3.2) as it can be expressed by the lower bound

$$E(N) \geq -\kappa N. \tag{3.3}$$

Put differently, the lowest possible energy calculated per particle cannot be arbitrarily negative as the number of particles increases. This is also often referred to as *stability of the second kind* [20, 21].

A maybe more intuitive notion of stability would be to ask for the volume occupied by a macroscopic object (in its ground state). Usually this volume is proportional to the number of particles N. Denoting by Ω a domain in $\mathbb{R}^3$ which is occupied by the system under consideration and by $E(\Omega)$ its (lowest possible) energy, (3.2) then reads

$$E(\Omega) \sim_{|\Omega|\to\infty} \bar{e}|\Omega| \tag{3.4}$$

where $|\Omega|$ is the volume of Ω. Stability of the second kind is expressed as

$$E(\Omega) \geq -\kappa|\Omega|. \tag{3.5}$$

Instead of the ground state energy, one can similarly consider the free energy $F(\Omega, \beta, \mu)$ at temperature $T = 1/\beta$ and chemical potential μ. One is then interested in proving the equivalent of (3.4)

$$F(\Omega, \beta, \mu) \sim_{|\Omega|\to\infty} \bar{f}(\beta, \mu)|\Omega| \tag{3.6}$$

where $\bar{f}(\beta, \mu)$ is the free energy per unit volume.

Large quantum Coulomb systems have been the object of an important investigation in the last decades and many techniques have been developed. A result like (3.3) (or equivalently (3.5)) was first proved for quantum electrons and nuclei by Dyson and Lenard [8]. After the original proof by Dyson and Lenard several other proofs were given. Lieb and Thirring [28] in particular presented an elegant and simple proof relying on an uncertainty principle for fermions. The different techniques and results concerning stability of matter were reviewed in several articles [20, 21, 22, 30, 33].

It is very important that the negatively charged particles (the electrons) are fermions. It was discovered by Dyson [7] that the Pauli exclusion principle is essential for Coulomb systems: charged bosons are alone not stable because their ground state energy satisfies $E(N) \sim -CN^{7/5}$, as was proved later [4, 27, 34].

A result like (3.2) (or equivalently (3.4)) was first proved by Lieb and Lebowitz [23] for a system containing electrons and nuclei both considered as quantum particles, hence invariant by rotation. Later Fefferman gave a different proof [9] for the case where the nuclei are classical particles placed on a lattice, a system which is not invariant by rotation.

In a recent work [16, 17, 15], we have provided a new insight in the study of the thermodynamic limit of quantum systems, by giving a general proof of (3.4) or (3.6) which can be applied to many different quantum systems including those studied by Lieb and Lebowitz [23] or Fefferman [9], and others which were not considered before. Our goal was to identify the main general physical properties of the free energy which are sufficient to prove the existence of the thermodynamic limit. However, for the sake of simplicity we will essentially address the crystal case in this paper and we refer to our works [16, 17] for a detailed study of the other cases.

In proving the existence of the thermodynamic limit of Coulomb quantum systems, the most difficult task is to quantify *screening*. Screening means that matter is arranged in such a way that it is essentially locally neutral, hence the electrostatic potential created by any subsystem decays much faster than expected. This effect is the main reason of the stability of ordinary matter but it is very subtle in the framework of quantum mechanics because the particles are by essence delocalized. In our approach, we shall heavily rely on an electrostatic inequality which was proved by Graf and Schenker [13, 12] and which serves as a way to quantify screening. It was itself inspired by previous works of Conlon, Lieb and Yau [4, 5], for systems interacting with the Yukawa potential. Fefferman used a similar idea in his study of the crystal case [9].

Like in previous works, our method consists in first showing the existence of the limit (3.6) for a specific domain $\triangle$ which is dilated (and possibly rotated and translated). Usually $\triangle$ is chosen to be a ball, a cube or a tetrahedron. In the applications [17] we always choose a tetrahedron as we shall use the Graf-Schenker inequality [13] which holds for this type of domains. The second step consists in showing the existence of the limit (3.6) for any (reasonable) sequence of domains $\{\Omega_n\}$ such that $|\Omega_n| \to \infty$. This is important as in principle the limit could depend on the chosen sequence, a fact that we want to exclude for our systems. We shall specify later what a "reasonable" sequence is. Essentially some properties will be needed to ensure that boundary effects always stay negligible.

It is to be noticed that our method (relying on the Graf-Schenker inequality) is primarily devoted to the study of quantum systems interacting through Coulomb forces. It might be applicable to other interactions but we shall not address this question here.

Proving a result like (3.4) or (3.6) is only a first step in the study of the thermodynamic limit of Coulomb quantum systems. An interesting open problem is to prove the convergence of *states* (or for instance of all k-body density matrices) and not only of energy levels. For the crystal case, convergence of the charge density or of the first order density matrix was proved for simplified models from Density Functional Theory or from Hartree-Fock theory [26, 3]. A result of this type was also proved for the Hartree-Fock approximation of no-photon Quantum Electrodynamics [14], as reviewed in Chapter 4.

Another (related) open question is to determine the next order in the asymptotics of the energy in the presence of local perturbations. Assume for instance that the crystal possesses a local defect modelled by a local potential V and denote the ground state energy in the domain Ω by $E^V(\Omega)$. Since V is local, it does not contribute to the energy in the first order of the thermodynamic limit. One is then interested in proving a behavior like $E^V(\Omega) = E^0(\Omega) + f(V) + o(1)_{|\Omega|\to\infty}$. Such a result was recently proved for the *reduced* Hartree-Fock model of the crystal with the exchange term neglected [2] as will be explained in Chapter 5. This includes an identification of the function $f(V)$. This program was also tackled for the Hartree-Fock model (with exchange term) of no-photon Quantum Electrodynamics [14], see Chapter 4.

The chapter is organized as follows. In the first section we introduce the model for the crystal and state our main theorem. In Section 2, we briefly describe two other quantum systems which we can treat using our method. Section 3 is devoted to the presentation of our new approach, in a quite general setting, together with hints on how it can be applied to the crystal case.

3.2 The crystal case

For simplicity, we put identical nuclei of charge +1 on each site of $\mathbb{Z}^3$. The results below can be generalized to any periodic system. Let Ω be a bounded open set of $\mathbb{R}^3$ and define the N-body Hamiltonian in Ω by

$$H_\Omega^N := \sum_{i=1}^N -\frac{\Delta_{x_i}}{2} + V_\Omega(x_1, ..., x_N),$$

where

$$V_\Omega(x) = \sum_{i=1}^N \sum_{R\in\mathbb{Z}^3\cap\Omega} \frac{-1}{|R-x_i|} + \frac{1}{2}\sum_{1\leq i\neq j\leq N}\frac{1}{|x_i-x_j|} + \frac{1}{2}\sum_{R\neq R'\in\mathbb{Z}^3\cap\Omega}\frac{1}{|R-R'|}.$$

Here $-\Delta$ is the *Dirichlet Laplacian* on Ω (we could as well consider another boundary condition). The Hamiltonian H_Ω^N acts on N-body fermionic wavefunctions $\Psi(x_1, .., x_N) \in \bigwedge_1^N L^2(\Omega)$. Stability of the first kind states that the spectrum of H_Ω^N is bounded from below:

$$E_\Omega^N = \inf_{\substack{\Psi\in\bigwedge_1^N H_0^1(\Omega),\\ \|\Psi\|_{L^2}=1}} \langle\Psi, H_\Omega^N\Psi\rangle = \inf\sigma_{\bigwedge_1^N L^2(\Omega)}(H_\Omega^N) > -\infty.$$

We may define the ground state energy in Ω by

$$E(\Omega) := \inf_{N\geq 0} E_\Omega^N. \tag{3.7}$$

It is more convenient to express (3.7) in a grand canonical formalism. We define the (electronic) Fock space as

$$\mathcal{F}_\Omega := \mathbb{C}\oplus\bigoplus_{N\geq1}\bigwedge_1^N L^2(\Omega)$$

The grand canonical Hamiltonian is then given by $H_\Omega := \bigoplus_{N\geq 0} H_\Omega^N$ with the convention that $H_\Omega^0 = (1/2)\sum_{R\neq R'\in\mathbb{Z}^3\cap\Omega}|R-R'|^{-1} \in \mathbb{R}\subset\mathbb{C}$. The number operator reads $\mathcal{N} := \bigoplus_{N\geq0} N$. It is then straightforward to check that

$$E(\Omega) = \inf\sigma_{\mathcal{F}_\Omega}(H_\Omega) = \inf_{\substack{\Gamma\in\mathcal{B}(\mathcal{F}_\Omega),\ \Gamma^*=\Gamma,\\ 0\leq\Gamma\leq1,\ \mathrm{Tr}_{\mathcal{F}_\Omega}(\Gamma)=1.}} \mathrm{Tr}_{\mathcal{F}_\Omega}(H_\Omega\Gamma).$$

The free energy at temperature $1/\beta$ and chemical potential $\mu\in\mathbb{R}$ is defined by

$$\begin{aligned} F(\Omega,\beta,\mu) &:= \inf_{\substack{\Gamma\in\mathcal{B}(\mathcal{F}_\Omega),\ \Gamma^*=\Gamma,\\ 0\leq\Gamma\leq1,\ \mathrm{Tr}_{\mathcal{F}_\Omega}(\Gamma)=1.}} \left(\mathrm{Tr}_{\mathcal{F}_\Omega}((H_\Omega-\mu\mathcal{N})\Gamma) + \frac{1}{\beta}\mathrm{Tr}_{\mathcal{F}_\Omega}(\Gamma\log\Gamma)\right) \\ &= -\frac{1}{\beta}\log\mathrm{Tr}_{\mathcal{F}_\Omega}\left[e^{-\beta(H_\Omega-\mu\mathcal{N})}\right]. \end{aligned} \tag{3.8}$$

As explained before, our purpose is to prove that

$$E(\Omega)\sim_{|\Omega|\to\infty}\bar{e}|\Omega| \quad\text{and}\quad F(\Omega,\beta,\mu)\sim_{|\Omega|\to\infty}\bar{f}(\beta,\mu)|\Omega| \tag{3.9}$$

in an appropriate sense. The first important property of E and F is the stability of matter.

Theorem 3.1 (Stability of Matter [17]). *There exists a constant C such that the following holds:*

$$E(\Omega)\geq -C|\Omega|, \qquad F(\Omega,\beta,\mu)\geq -C\left(1+\beta^{-5/2}+\max(0,\mu)^{5/2}\right)|\Omega|$$

for any bounded open set $\Omega\subset\mathbb{R}^3$ and any $\beta>0$, $\mu\in\mathbb{R}$.

Sketch of the proof. The first step is to use an inequality for classical systems due to Baxter [1], improved later by Lieb and Yau [29], and which allows to bound the full N-body Coulomb potential by a one-body potential:

$$V(x_1,...,x_N)\geq -\sum_{i=1}^N\frac{3/2+\sqrt{2}}{\delta(x_i)} \tag{3.10}$$

where $\delta(x)=\inf_{R\in\mathbb{Z}^3}|x-R|$ is the distance to the closest nucleus. Hence we have the lower bound

$$H^N_\Omega\geq\sum_{i=1}^N\left(-\frac{\Delta_{x_i}}{2}-\frac{3/2+\sqrt{2}}{\delta(x_i)}\right).$$

Next we split the kinetic energy in two parts and we use the uncertainty principle to show that on $L^2(\Omega)$

$$-\frac{\Delta}{4}-\frac{3/2+\sqrt{2}}{\delta(x)}\geq -C.$$

In proving this lower bound, one uses the Sobolev inequality in a small ball around each nucleus, exploiting the fact that the nuclei are fixed and separated by a distance at least one to each other. The proof of the stability of matter for systems with classical nuclei whose position is unknown is more difficult and it uses the improved version of (3.10) contained in the paper by Lieb and Yau [29], as explained in our work [17]. This shows

$$H^N_\Omega\geq\sum_{i=1}^N\left(-\frac{\Delta_{x_i}}{4}-C\right)\quad\text{hence}\quad H_\Omega\geq-\frac{1}{4}\sum_i\Delta_i-C\mathcal{N} \tag{3.11}$$

on $\bigwedge_1^N L^2(\Omega)$ and $\mathcal{F}_\Omega$ respectively. The last step is to use the Lieb-Thirring inequality [28] which states that

$$\left\langle\sum_{i=1}^N(-\Delta_{x_i})\,\Psi,\Psi\right\rangle\geq C_{\rm LT}\int_\Omega\rho_\Psi(x)^{5/3}dx \tag{3.12}$$

for all $N\geq 1$ and all N-body fermionic wavefunction $\Psi\in\bigwedge_1^N L^2(\Omega)$. The density of charge ρ_Ψ is as usual defined by $\rho_\Psi(x)=N\int_{\Omega^{N-1}}|\Psi(x,y)|^2dy$. Using the fact that $\int_\Omega\rho_\Psi=N$ and Hölder's inequality, (3.12) yields on the Fock space $\mathcal{F}_\Omega$

$$\sum_i(-\Delta_{x_i})\geq C_{\rm LT}|\Omega|^{-2/3}\mathcal{N}^{5/3}. \tag{3.13}$$

Hence we obtain $H_\Omega\geq(C_{\rm LT}/4)|\Omega|^{-2/3}\mathcal{N}^{5/3}-C\mathcal{N}$ which, when optimized over N, gives the result for the ground state energy.

For the free energy, we use (3.11), (3.13) and Peierls' inequality [32, 35] to get

$$F(\beta,\mu,\Omega)\geq-\frac{1}{\beta}\log{\rm Tr}_{\mathcal{F}}\left(e^{-\beta\sum_i(-\Delta_i)/4}\right)-C(1+\mu_+^{5/2})|\Omega|.$$

The first term of the r.h.s. is the free energy of a free-electron gas which is bounded below by $-C(1+\beta^{-5/2})|\Omega|$ in the thermodynamic limit [17]. □

In order to state our main result, we need the following

Definition 3.1 (Regular sets in $\mathbb{R}^3$)**.** *Let $a>0$ and $\epsilon>0$. We say that a bounded open set $\Omega\subseteq\mathbb{R}^3$ has* an a-regular boundary in the sense of Fisher *if, denoting by $\partial\Omega=\overline{\Omega}\setminus\Omega$ the boundary of Ω,*

$$\forall t\in[0,1],\qquad\left|\left\{x\in\mathbb{R}^3\ |\ {\rm d}(x,\partial\Omega)\leq|\Omega|^{1/3}t\right\}\right|\leq|\Omega|\,a\,t. \tag{3.14}$$

We say that a bounded open set $\Omega \subseteq \mathbb{R}^3$ *satisfies the* ε-cone property *if for any* $x \in \Omega$ *there is a unit vector* $a_x \in \mathbb{R}^3$ *such that*

$$\{y \in \mathbb{R}^3 \mid (x-y)\cdot a_x > (1-\varepsilon^2)|x-y|,\ |x-y| < \varepsilon\} \subseteq \Omega.$$

We denote by $\mathcal{R}_{a,\varepsilon}$ *the set of all* $\Omega \subseteq \mathbb{R}^3$ *which have an* a*-regular boundary and such that both* Ω *and* $\mathbb{R}^3 \setminus \Omega$ *satisfy the* ε*-cone property.*

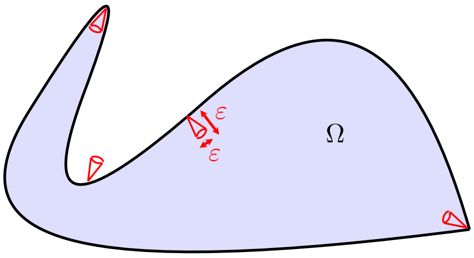


Figure 3.1: Cone property.

Note that any open convex set is in $\mathcal{R}_{a,\varepsilon}$ for some $a > 0$ large enough and $\varepsilon > 0$ small enough [16]. We may state our main

Theorem 3.2 (Thermodynamic Limit for the Crystal [17])**.** *There exist* $\bar{e} \in \mathbb{R}$ *and a function* $\bar{f} : (0,\infty)\times\mathbb{R} \to \mathbb{R}$ *such that the following holds: for any sequence* $\{\Omega_n\}_{n\geq 1} \subseteq \mathcal{R}_{a,\epsilon}$ *of domains with* $|\Omega_n| \to \infty$, $|\Omega_n|^{-1/3}\mathrm{diam}(\Omega_n) \leq C$, $a \geq a_0 > 0$ *and* $0 < \varepsilon \leq \varepsilon_0$

$$\lim_{n\to\infty} \frac{E(\Omega_n)}{|\Omega_n|} = \bar{e}, \qquad \lim_{n\to\infty} \frac{F(\Omega_n,\beta,\mu)}{|\Omega_n|} = \bar{f}(\beta,\mu). \tag{3.15}$$

Moreover $\bar{f}$ *takes the form* $\bar{f}(\beta,\mu) = \varphi(\beta) - \mu$.

Remark 3.1. *We know from [23, Appendix A p. 385] and [10, Lemma 1] that if each set* Ω_n *of the considered sequence is connected, then automatically* $|\Omega_n|^{-1/3}\mathrm{diam}(\Omega_n) \leq C$.

A very similar result was proved by C. Fefferman [9]. Our result is more general: we allow any sequence Ω_n tending to infinity and which is regular in the sense that $\{\Omega_n\}_{n\geq 1} \subseteq \mathcal{R}_{a,\epsilon}$. In Fefferman's paper [9], $\Omega_n = \ell_n(\Omega + x_n)$ where $\ell_n \to \infty$, Ω is a fixed convex open set and x_n is any sequence in $\mathbb{R}^3$. These sets are always in $\mathcal{R}_{a,\epsilon}$ for some $a, \varepsilon > 0$.

In our work [17] a result even more general than Theorem 3.2 is shown: we are able to prove the existence of the same thermodynamic limit if the crystal is *locally perturbed* (for instance finitely many nuclei are moved or their charge is changed). A similar result can also be proved for the Hartree-Fock model.

3.3 Other models

Our approach [16, 17] is general and it can be applied to a variety of models, not only the crystal case. We quickly mention two such examples. It is interesting to note that for these other models, we do not need the cone property and we can weaken the assumptions on the regularity of the boundary by replacing t on the r.h.s. of (3.14) by any t^p, $0 < p \leq 1$. Details may be found in our article [17]. Roughly speaking, when the system is “rigid” like for the crystal (the nuclei are fixed), the proof is more complicated and more assumptions are needed on the sequence of domains to avoid undesirable boundary effects.

3.3.1 Quantum particles in a periodic magnetic field.

Define the magnetic kinetic energy $T(A) = (-i\nabla + A(x))^2$ where $B = \nabla \times A$ is periodic (for instance constant) and $A \in L^2_{\rm loc}(\mathbb{R}^3)$. Next, consider the Hamiltonian

$$H_\Omega^{N,K} := \sum_{i=1}^N T(A)_{x_i} + \sum_{k=1}^K T(A)_{R_k} + V(x,R),$$

$$V(x,R) = \sum_{i,k} \frac{-z}{|R_k - x_i|} + \frac{1}{2}\sum_{i\neq j} \frac{1}{|x_i - x_j|} + \frac{1}{2}\sum_{k\neq k'} \frac{z^2}{|R_k - R_{k'}|}$$

The ground state energy is this time defined as

$$E'(\Omega) := \inf_{N,K\geq 0} \inf \sigma_{\bigwedge_1^N L^2(\Omega)\otimes S \bigotimes_1^K L^2(\Omega)} \left(H_\Omega^{N,K}\right).$$

We do not precise the symmetry S of the particles of charge z which can be bosons or fermions. A formula similar to (3.8) may be used for the free energy on the (electronic and nucleic) Fock space. We prove in our paper [17] a result similar to Theorem 3.2 for this model. Lieb and Lebowitz already proved it in the seminal paper [23] when $A \equiv 0$. They used as an essential tool the rotation-invariance of the system to obtain screening. When $A \neq 0$ the system is no more invariant by rotations and their method cannot be applied.

3.3.2 Classical nuclei with optimized position.

For all $R \subset \Omega$, $\#R < \infty$, let us define

$$H_\Omega^{N,R} := \sum_{i=1}^N -\frac{\Delta_{x_i}}{2} + V(x,R)$$

and the associated ground state energy by

$$E''(\Omega) := \inf_{N\geq 0} \inf_{\substack{R\subset\Omega,\\ \#R<\infty}} \inf \sigma_{\bigwedge_1^N L^2(\mathbb{R}^3)} \left(H_\Omega^{N,R}\right).$$

We could as well optimize the charges in $[0,z]$ of the nuclei without changing the energy [6, 17]. However, the free energy itself is not the same when the charges of the nuclei are optimized or not [17].

Surprisingly, to our knowledge the existence of the thermodynamic limit for this model was unknown. A result similar to Theorem 3.2 is proved in our paper [17] for E''.

3.4 A general method

In this section, we give the main ideas of our new approach which allows to prove Theorem 3.2 and its counterparts for the other models quoted before.

3.4.1 Screening via the Graf-Schenker inequality

As mentioned in the introduction, an important step is to quantify screening. For quantum nuclei without a magnetic field ($A \equiv 0$), Lieb and Lebowitz used [23] the following method (see Figure 3.2). First they took a big ball B which they packed with several small balls B_k of different size. In each of these balls, they took the (neutral) ground state of the corresponding ball. As the system is invariant under rotations, they can freely rotate each ground state. Averaging over rotations of all

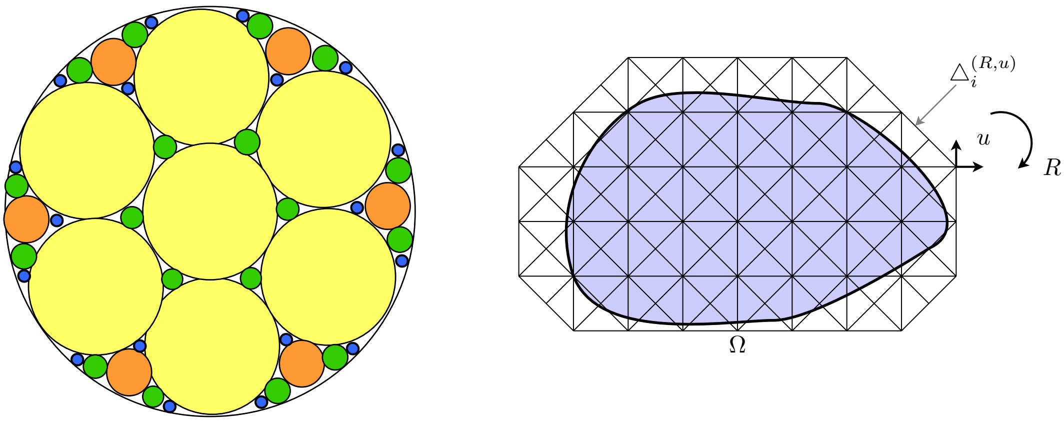


Figure 3.2: A comparison between the original method of Lieb and Lebowitz [23] (left) and our method based on the Graf-Schenker inequality [13, 16, 17] (right).

the small balls, they reduced the computation of the interaction between them to that of classical pointwise particles located at the center of the balls, by Newton's theorem. As each subsystem is neutral, this interaction vanishes. This proves *a fortiori* that there exists an adequate rotation of each system in each little ball such that the total interaction between them cancels. Choosing this configuration, they could build a test function whose energy is just the sum of the small energies, proving an estimate of the form $E(B) \leq \sum_k E(B_k)$. This inequality can be used to prove the limit for balls. Clearly this trick can only be used for rotation-invariant systems.

Note in the Lieb-Lebowitz proof, a domain (the big ball) is split in several *fixed subdomains* and an average is done *over rotations of the states* in each small domain. This yields an *upper bound* to the energy. The Graf-Schenker inequality is kind of dual to the above method (see Figure 3.2). This time a domain Ω is split in several subdomains by using a tiling of the space $\mathbb{R}^3$. But the system is frozen in the state of the big domain Ω and the average is done *over the position of the tiling*. This yields a *lower bound* to the energy of the form $E(\Omega) \geq \sum_k E(\Delta_i^{(R,u)} \cap \Omega) + \text{errors}$, where $\Delta_i^{(R,u)}$ are the tetrahedrons which make up the (translated and rotated) tiling.

The Graf-Schenker inequality was inspired by previous works of Conlon, Lieb and Yau [4, 5]. It is an estimate on the Coulomb energy of classical particles. The proof of Fefferman in the crystal case [9] was also based on a *lower bound* on the free energy in a big set and an average over translations of a covering of this set (the method was reexplained later in details by Hugues [18]). Fefferman [9] uses a covering with balls and cubes of different size. The lower bound depends on the number of balls contained in the big domain and of the form of the kinetic energy which is used to control error terms.

Let $G = \mathbb{R}^3 \rtimes SO_3(\mathbb{R})$ be the group of translations and rotations acting on $\mathbb{R}^3$, and denote by $d\lambda(g)$ its Haar measure.

Lemma 3.1 (Graf-Schenker inequality [13])**.** *Let $\triangle$ be a simplex in $\mathbb{R}^3$. There exists a constant C such that for any $N \in \mathbb{N}$, $z_1, ..., z_N \in \mathbb{R}$, $x_i \in \mathbb{R}^3$ and any $\ell > 0$,*

$$\sum_{1 \leq i < j \leq N} \frac{z_i z_j}{|x_i - x_j|} \geq \int_G \frac{d\lambda(g)}{|\ell\triangle|} \sum_{1 \leq i < j \leq N} \frac{z_i z_j \mathbb{1}_{g\ell\triangle}(x_i) \mathbb{1}_{g\ell\triangle}(x_j)}{|x_i - x_j|} - \frac{C}{\ell} \sum_{i=1}^N z_i^2. \tag{3.16}$$

In the previous lemma it is not assumed that $\triangle$ yields a tiling of $\mathbb{R}^3$. Up to an error which scales like ℓ, (3.16) says that the total Coulomb energy can be bounded from below by the Coulomb energy (per unit volume) of the particles which are in the (dilated) simplex $g\ell\triangle$, averaged over all translations and rotations g of this simplex.

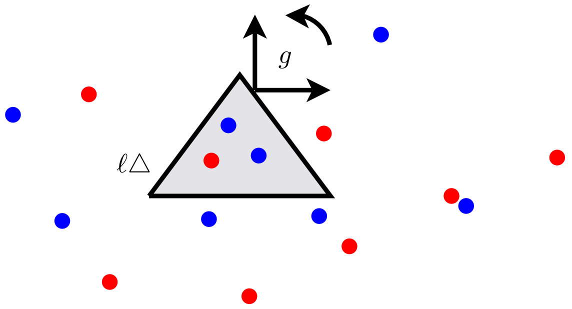


Figure 3.3: The Graf-Schenker inequality: the total Coulomb interaction of the charged particles is bounded below by an average over translations and rotations of a term corresponding to the Coulomb interaction of the particles which are in the simplex only, plus an error term.

Because of the above inequality, simplices play a specific role in the study of Coulomb systems. Hence proving the existence of the thermodynamic limit for simplices first is natural (as it was natural to consider balls in the Lieb-Lebowitz case due to the invariance by rotation). In the next section we give an abstract setting for proving the existence of the limit when an inequality of the form (3.16) holds true.

It is to be noticed that, so far, an inequality of the type (3.16) is only known in *dimension 3* for the *Coulomb potential* (it can easily be generalized in any dimension but always for the potential $1/|x|$ which is not the Coulomb one in dimension $d \neq 3$) and for *simplices* (the inequality can be shown to be wrong for balls). A generalization to other interactions and other convex domains (with possibly a different error term) is a very interesting question.

3.4.2 An abstract result

In this section we consider an abstract energy $E : \Omega \in \mathcal{M} \mapsto E(\Omega) \in \mathbb{R}$ defined on the set $\mathcal{M}$ of all bounded open subsets of $\mathbb{R}^3$ and we give sufficient conditions for the existence of the thermodynamic limit. In the application, E will be either the ground state energy, or the free energy of the system under consideration.

We fix a reference set $\triangle \in \mathcal{R}_{a,\epsilon}$ which is only assumed to be a bounded open convex set in $\mathbb{R}^3$ (it need not be a simplex for this section), such that $0 \in \triangle$. Here $a, \epsilon > 0$ are fixed. We assume that the energy E satisfies the following five assumptions:

(A1) *(Normalization).* $E(\emptyset) = 0$.

(A2) *(Stability).* $\forall \Omega \in \mathcal{M}$, $E(\Omega) \geq -\kappa|\Omega|$.

(A3) *(Translation Invariance).* $\forall \Omega \in \mathcal{R}_{a,\epsilon}$, $\forall z \in \mathbb{Z}^3$, $E(\Omega + z) = E(\Omega)$.

(A4) *(Continuity).* $\forall \Omega \in \mathcal{R}_{a,\epsilon}, \Omega' \in \mathcal{R}_{a',\epsilon'}$ with $\Omega' \subseteq \Omega$ and $\mathrm{d}(\partial\Omega, \partial\Omega') > \delta$,

$$E(\Omega) \leq E(\Omega') + \kappa|\Omega \setminus \Omega'| + |\Omega|\alpha(|\Omega|).$$

(A5) *(Subaverage Property).* For all $\Omega \in \mathcal{M}$, we have

$$E(\Omega) \geq \frac{1-\alpha(\ell)}{|\ell\triangle|} \int_G E\big(\Omega \cap g\cdot(\ell\triangle)\big)\, d\lambda(g) - |\Omega|_{\mathrm{r}}\, \alpha(\ell) \tag{3.17}$$

where $|\Omega|_{\mathrm{r}} := \inf\{|\tilde{\Omega}|,\ \Omega \subseteq \tilde{\Omega},\ \tilde{\Omega} \in \mathcal{R}_{a,\epsilon}\}$ is the regularized volume of Ω.

In the assumptions above α is a fixed function which tends to 0 at infinity and δ, a', ϵ' are fixed positive constants. In our work [16], an even more general setting is provided. First **(A3)** can

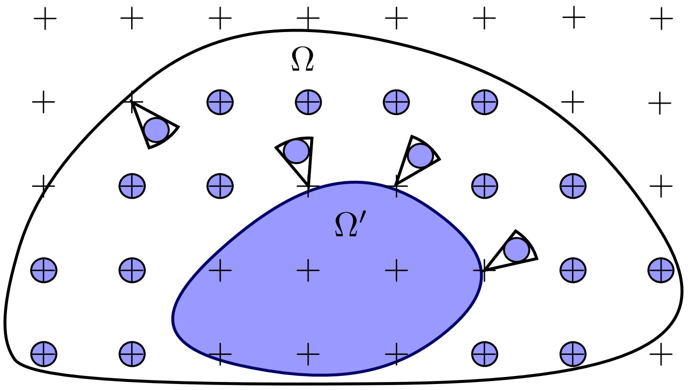


Figure 3.4: Idea of the proof of **(A4)** for the crystal.

be replaced by a much weaker assumption but we do not detail this here. Also a generic class of regular sets $\mathcal{R}$ is considered instead of $\mathcal{R}_{a,\epsilon}$. This is because for instance the cone property is only needed for the crystal case and it is not at all necessary in other models, hence the concept of regularity depends on the application.

Notice **(A4)** essentially says that a small decrease of Ω will not decrease too much the energy. A similar property was used and proved in the crystal case by Fefferman [9, Lemma 2]. Taking $\Omega' = \emptyset$ and using **(A1)**, property **(A4)** in particular implies that for any regular set $\Omega \in \mathcal{R}_{a,\epsilon}$, $E(\Omega) \leq C|\Omega|$. However this upper bound need not be true for all $\Omega \in \mathcal{M}$. We give a sketch of the proof of the following result in Section 3.4.5.

Theorem 3.3 (Abstract Thermodynamic Limit for $\triangle$ [16])**.** *Assume $E : \mathcal{M} \to \mathbb{R}$ satisfies the above properties **(A1)–(A5)** for some open convex set $\triangle \in \mathcal{R}_{a,\epsilon}$ with $0 \in \triangle$. There exists $\bar{e} \in \mathbb{R}$ such that $e_\ell(g) = |\ell\triangle|^{-1}E\big(g\ell\triangle\big)$ converges uniformly towards $\bar{e}$ for $g \in G = \mathbb{R}^3 \rtimes SO(3)$ and as $\ell \to \infty$. Additionally, the limit $\bar{e}$ does not depend on the set $\triangle$[1].*

3.4.3 Idea of the proof of (A1)–(A5) for the crystal

Before switching to the abstract case of a general sequence $\{\Omega_n\}$, we give an idea of the proof of **(A1)**–**(A5)** in the crystal case. We apply the theory of the previous section to both the ground state energy and the free energy of the crystal which were defined in Section 3.2. First **(A1)** and **(A3)** are obvious. Property **(A2)** is the stability of matter as stated in Theorem 3.1. On the other hand **(A5)** is essentially the Graf-Schenker inequality (3.16), up to some localization issues of the kinetic energy which have essentially already been dealt with by Graf and Schenker [13].

For the crystal the most difficult property is **(A4)**. The difficulty arises from the fact that this is a very rigid system. For the two other examples mentioned in Section 3.3, **(A4)** is obvious, the energy being nonincreasing: $E(\Omega) \leq E(\Omega')$. This is because we can simply choose a ground state of Ω' as a test for Ω and take the vacuum in $\Omega \setminus \Omega'$. In the crystal case we always have nuclei in $\Omega \setminus \Omega'$ and if we do not put any electron to screen them, they will create an enormous electrostatic energy.

The idea of the proof of **(A4)** for the crystal is displayed in Figure 3.4. We build a test state in Ω by considering the ground state in Ω', and placing one radial electron in a ball of fixed size on top of each nucleus outside Ω'. By Newton's theorem, the electrostatic potential out of the support of the electron will vanish, hence the energy will simply be $E(\Omega')$ plus the sum of the kinetic energies of the electrons, which is bounded above by a constant times $|\Omega \setminus \Omega'|$ for regular domains. The only problem is that we cannot put an electron on top of the nuclei which are too close to the boundary of Ω or of Ω'. For these nuclei, using the cone property we can place the ball

---

[1] This means if all the assumptions are true for another set $\triangle'$ then one must have $\bar{e}' = \bar{e}$.

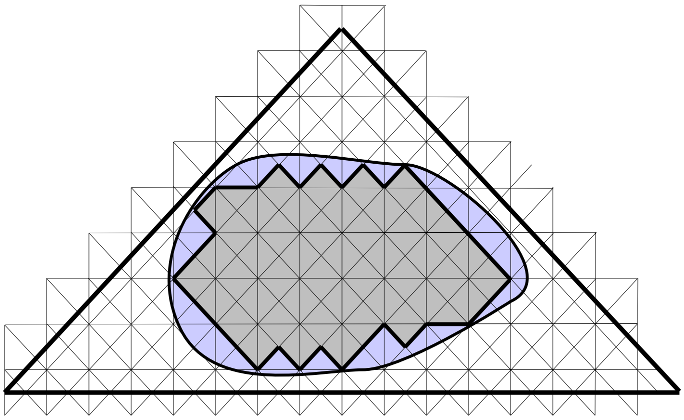


Figure 3.5: Proof for general sequences $\{\Omega_n\}$.

aside and create a dipole. The difficult task is then to compute a bound on the total interaction between the dipoles and the ground state in Ω'. We prove [17] that it is $o(|\Omega|)$, using a specific version of stability of matter.

3.4.4 General domains and strong subadditivity of entropy

In the previous two subsections, we have presented our abstract theory giving the thermodynamic limit of special sequences built upon the reference set $\triangle$, and we have explained how to apply it to the crystal case. For all regular domain sequences we can only get from **(A5)** a bound of the form

$$\liminf_{n\to\infty} \frac{E(\Omega_n)}{|\Omega_n|} \geq \bar{e}.$$

In order to get the upper bound, we use a big simplex $L_n\triangle$ of the same size as Ω_n and a tiling made with simplices of size $\ell_n \ll L_n$, as shown in Figure 3.5. We use the ground state of the big simplex $L_n\triangle$ to build a test state in Ω_n, hence giving the appropriate upper bound. To this end, we need some localization features, hence more assumptions in the general theory.

It is sufficient [16] to assume that

(i) $\triangle$ can be used to build a tiling of $\mathbb{R}^3$;

(ii) the free energy is essentially "two-body"[2] such that we may write the total energy $E(L_n\triangle)$ as the sum of the energies of the small sets of the tiling, plus the interaction between them and the relative entropy;

(iii) the entropy is strongly subadditive.

This is summarized in the following assumption. We assume that Γ is a subgroup of G yielding a tiling of $\mathbb{R}^3$ by means of $\triangle$, i.e. $\overline{\cup_{\mu\in\Gamma}\mu\triangle} = \mathbb{R}^3$ and $\mu\triangle \cap \nu\triangle = \emptyset$ for $\mu \neq \nu$.

(A6) *(Two-body decomposition).* For all L and ℓ we can find $g \in G$ and maps $E_g : \Gamma \to \mathbb{R}$, $I_g : \Gamma \times \Gamma \to \mathbb{R}$, $s_g : \{\mathcal{P} : \mathcal{P} \subseteq \Gamma\} \to \mathbb{R}$ such that

- $E_g(\mu) = I_g(\mu, \nu) = 0$ if $\ell g\mu\triangle \cap (L\triangle) = \emptyset$;
- $E(L\triangle) \geq \sum_{\mu\in\Gamma} E_g(\mu) + \frac{1}{2} \sum_{\substack{\mu,\nu\in\Gamma\\ \mu\neq\nu}} I_g(\mu,\nu) - s_g(\Gamma) - |L\triangle|\alpha(\ell);$

---

[2] We could as well assume that the energy is k-body with $k < \infty$ but this would complicate the assumptions even more.

- For all $\mathcal{P} \subseteq \Gamma$ and $A_{\mathcal{P}} = L\triangle \cap \bigcup_{\mu\in\mathcal{P}} \ell g\mu\triangle$

$$E(A_{\mathcal{P}}) \leq \sum_{\mu\in\mathcal{P}} E_g(\mu) + \frac{1}{2} \sum_{\substack{\mu,\nu\in\mathcal{P}\\ \mu\neq\nu}} I_g(\mu,\nu) - s_g(\mathcal{P}) + |A_{\mathcal{P}}|\alpha(\ell);$$

- *(Strong subadditivity).* for any disjoint subsets $\mathcal{P}_1$, $\mathcal{P}_2$, $\mathcal{P}_3 \subseteq \Gamma$

$$s_g(\mathcal{P}_1 \cup \mathcal{P}_2 \cup \mathcal{P}_3) + s_g(\mathcal{P}_2) \leq s_g(\mathcal{P}_1 \cup \mathcal{P}_2) + s_g(\mathcal{P}_2 \cup \mathcal{P}_3);$$

- *(Subaverage property).* $\displaystyle\int_{G/\Gamma} dg \sum_{\substack{\mu,\nu\in\Gamma\\ \mu\neq\nu}} I_g(\mu,\nu) \geq -|L\triangle|\alpha(\ell)$.

In the applications[3] the previous quantities are interpreted as follows: $E_g(\mathcal{P})$ is the free energy in the union $A_{\mathcal{P}} = (L\triangle)\cap\cup_{\mu\in\mathcal{P}}\ell g\mu\triangle$, $I_g(\mu,\nu)$ is the interaction energy between the simplices $\ell g\mu\triangle$ and $\ell g\nu\triangle$, and $s_g(\mathcal{P})$ is the difference between the entropy of $A_{\mathcal{P}}$ and the sum of the entropies of $\ell g\mu\triangle$ with $\mu \in \mathcal{P}$.

Conjectured by Lanford and Robinson [19] the strong subadditivity (SSA) of the entropy in the quantum mechanical case was proved by Lieb and Ruskai [24, 25]. The fact that SSA is very important in the thermodynamic limit was remarked by Robinson and Ruelle [31] and others [35]. In [16] we prove the following

Theorem 3.4 (Abstract Limit for general domains [16])**.** *Assume $E : \mathcal{M} \to \mathbb{R}$ satisfies the properties **(A1)–(A6)** for some open convex polyhedron $\triangle \in \mathcal{R}_{a,\epsilon}$ with $0 \in \triangle$, yielding a tiling of $\mathbb{R}^3$. Then we have for all sequences $\{\Omega_n\} \subset \mathcal{R}_{a,\epsilon}$ with $|\Omega_n| \to \infty$ and $|\Omega_n|^{-1/3}\mathrm{diam}(\Omega_n) \leq C$,*

$$\lim_{n\to\infty} \frac{E(\Omega_n)}{|\Omega_n|} = \bar{e}$$

where $\bar{e}$ is the limit obtained in Theorem 3.3.

The proof of Theorem 3.4 is based on a careful estimate of the energy and the interaction energies of boundary terms, ie. of the sets $\ell g\mu\triangle$ which intersect the boundary of the big set $L\triangle$. The application to the crystal is not much more difficult than for Theorem 3.3. Indeed in the paper of Graf and Schenker [13], (3.16) was expressed using a tiling of $\mathbb{R}^3$ and the last subaverage property of **(A6)** essentially follows from their ideas [13]. Strong subadditivity of the entropy is usually expressed via partial traces. A generalization in the setting of localization in Fock space is detailed in our article [17].

3.4.5 Proof of Theorem 3.3

Denote as in the Theorem $e_\ell(g) = E(g\ell\triangle)|\ell\triangle|^{-1}$. Notice that **(A2)**, **(A4)** with $\Omega' = \emptyset$, and **(A1)** imply that e_ℓ is uniformly bounded on G. Also we have by **(A3)** $e_\ell(u+z,R) = e_\ell(u,R)$ for all $(u,R) \in \mathbb{R}^3 \times SO_3(\mathbb{R})$, $z \in \mathbb{Z}^3$, i.e. e_ℓ is periodic with respect to translations. Hence it suffices to prove the theorem for $g = (u,R) \in [0,1]^3 \times SO_3(\mathbb{R})$.

Next we take $\bar{g} \in G$, $L \gg \ell$ and apply **(A5)** with $\Omega = \bar{g}L\triangle$. We get

$$e_L(\bar{g}) \geq \frac{1-\alpha(\ell)}{|L\triangle|} \int_G \frac{E(\bar{g}L\triangle \cap g\ell\triangle)}{|\ell\triangle|} dg - \alpha(\ell).$$

---

[3] Due to some localization issues of the kinetic energy, it is often needed that the sets of the tiling slightly overlap. See [16] for a generalization in this direction.

Let us introduce the set $\mathcal{Z}$ of points $z \in \mathbb{Z}^3$ such that $R\ell\triangle + u + z \subset \bar{g}L\triangle$ for all $u \in [0,1]^3$ and all $R \in SO_3(\mathbb{R})$. We also define $\partial\mathcal{Z}$ as the set of points $z \in \mathbb{Z}^3$ such that $(R\ell\triangle + u + z) \cap \bar{g}L\triangle \neq \emptyset$ for some $(u, R) \in [0,1]^3 \times SO_3(\mathbb{R})$ but $z \notin \mathcal{Z}$. We obtain using **(A1)** and **(A3)**

$$\begin{aligned}
\int_G \frac{E(\bar{g}L\triangle \cap g\ell\triangle)}{|\ell\triangle|}\,dg &= \sum_{z\in\mathbb{Z}^3} \int_{[0,1]^3} du \int_{SO_3(\mathbb{R})} dR \frac{E(\bar{g}L\triangle \cap (R\ell\triangle + u + z))}{|\ell\triangle|} \\
&= \sum_{z\in\partial\mathcal{Z}} \int_{[0,1]^3} du \int_{SO_3(\mathbb{R})} dR \frac{E(\bar{g}L\triangle \cap (R\ell\triangle + u + z))}{|\ell\triangle|} \\
&\qquad + (\#\mathcal{Z}) \int_{[0,1]^3} du \int_{SO_3(\mathbb{R})} dR\; e_\ell(u, R).
\end{aligned}$$

Using the stability property **(A2)**, we infer

$$\frac{E(\bar{g}L\triangle \cap (R\ell\triangle + u + z))}{|\ell\triangle|} \geq -\kappa \frac{|\bar{g}L\triangle \cap (R\ell\triangle + u + z)|}{|\ell\triangle|} \geq -\kappa.$$

Hence

$$\int_G \frac{E(\bar{g}L\triangle \cap g\ell\triangle)}{|\ell\triangle|}\,dg \geq (\#\mathcal{Z}) \int_{[0,1]^3 \times SO_3(\mathbb{R})} e_\ell(g)\,dg + \kappa(\#\partial\mathcal{Z}).$$

As $\triangle$ has an a-regular boundary, it can be seen that $(\#\partial\mathcal{Z}) \leq CL^2\ell$ and $\#\mathcal{Z} = |L\triangle| + O(L^2\ell)$. Using again that e_ℓ is bounded, we eventually obtain the estimate

$$e_L(\bar{g}) \geq \int_{[0,1]^3 \times SO_3(\mathbb{R})} e_\ell(g)\,dg - C(\alpha(\ell) + \ell/L)$$

for some constant C. It is then an easy exercise to prove that

$$\lim_{\ell\to\infty} \inf_G e_\ell = \lim_{\ell\to\infty} \int_{[0,1]^3 \times SO_3(\mathbb{R})} e_\ell := \bar{e}$$

and finally that $e_\ell \to \bar{e}$ in $L^1([0,1]^3 \times SO_3(\mathbb{R}))$.

The last step consists in proving the uniform convergence, using **(A4)**. Fix some small $\eta > 0$. As $0 \in \triangle$ and $\triangle$ is convex, we have $(1-\eta)\triangle \subset \triangle$. More precisely, there exists an $r > 0$ and a neighborhood W of the identity in $SO_3(\mathbb{R})$ such that $R(1-\eta)\triangle + u \subset \triangle$ for all $(u, R) \in A := B(0,r) \times W \subset G$. We have that $g\ell(1-\eta)\triangle \subset \ell\triangle$ for all $g \in A_\ell := B(0, r\ell) \times W$, hence in particular for all $g \in A$. Now we fix some $\bar{g} \in G$ and apply **(A4)** with $\Omega = \bar{g}\ell\triangle$ and $\Omega' = \bar{g}g\ell(1-\eta)\triangle$, we get

$$E(\bar{g}\ell\triangle) \leq E(\bar{g}g\ell(1-\eta)\triangle) + C|\ell\triangle|\eta + o(|\ell\triangle|).$$

Integrating over $g \in A$ and dividing by $|\ell\triangle|$ we infer

$$e_\ell(\bar{g}) \leq \frac{1}{|\bar{g}A|} \int_{\bar{g}A} e_{(1-\eta)\ell}(g)\,dg + C\eta + o(1)_{\ell\to\infty}.$$

First we pass to the limit as $\ell \to \infty$ using that $e_\ell \to \bar{e}$ in $L^1(G)$ and $|A| \neq 0$. Then we take $\eta \to 0$ and get $\limsup_{\ell\to\infty} \sup_{\bar{g}\in G} e_\ell(\bar{g}) \leq \bar{e}$. This ends the proof of Theorem 3.3. ☐

References

[1] J. R. Baxter, *Inequalities for potentials of particle systems*, Illinois J. Math., 24 (1980), pp. 645–652.

[2] É. Cancès, A. Deleurence, and M. Lewin, *A new approach to the modelling of local defects in crystals: the reduced Hartree-Fock case*, Commun. Math. Phys., 281 (2008), pp. 129–177.

[3] I. Catto, C. Le Bris, and P.-L. Lions, *On the thermodynamic limit for Hartree-Fock type models*, Ann. Inst. H. Poincaré Anal. Non Linéaire, 18 (2001), pp. 687–760.

[4] J. G. Conlon, E. H. Lieb, and H.-T. Yau, *The $N^{7/5}$ law for charged bosons*, Commun. Math. Phys., 116 (1988), pp. 417–448.

[5] ———, *The Coulomb gas at low temperature and low density*, Commun. Math. Phys., 125 (1989), pp. 153–180.

[6] I. Daubechies and E. H. Lieb, *One-electron relativistic molecules with Coulomb interaction*, Commun. Math. Phys., 90 (1983), pp. 497–510.

[7] F. J. Dyson, *Ground-state energy of a finite system of charged particles*, J. Math. Phys., 8 (1967), pp. 1538–1545.

[8] F. J. Dyson and A. Lenard, *Stability of matter. I*, J. Math. Phys., 8 (1967), pp. 423–434.

[9] C. Fefferman, *The thermodynamic limit for a crystal*, Commun. Math. Phys., 98 (1985), pp. 289–311.

[10] M. E. Fisher, *The free energy of a macroscopic system*, Arch. Rational Mech. Anal., 17 (1964), pp. 377–410.

[11] M. E. Fisher and D. Ruelle, *The stability of many-particle systems*, J. Math. Phys., 7 (1966), pp. 260–270.

[12] G. M. Graf, *Stability of matter through an electrostatic inequality*, Helv. Phys. Acta, 70 (1997), pp. 72–79. Papers honouring the 60th birthday of Klaus Hepp and of Walter Hunziker, Part II (Zürich, 1995).

[13] G. M. Graf and D. Schenker, *On the molecular limit of Coulomb gases*, Commun. Math. Phys., 174 (1995), pp. 215–227.

[14] C. Hainzl, M. Lewin, and J. P. Solovej, *The mean-field approximation in quantum electrodynamics: the no-photon case*, Comm. Pure Appl. Math., 60 (2007), pp. 546–596.

[15] ———, *The thermodynamic limit of quantum Coulomb systems: A new approach*, in Mathematical results in Quantum Mechanics: Proceedings of the QMath10 Conference, I. Beltita, G. Nenciu, and R. Purice, eds., World Scientific, 2008.

[16] ———, *The thermodynamic limit of quantum Coulomb systems. Part I. General theory*, Advances in Math., 221 (2009), pp. 454–487.

[17] ———, *The thermodynamic limit of quantum Coulomb systems. Part II. Applications*, Advances in Math., 221 (2009), pp. 488–546.

[18] W. Hughes, *Thermodynamics for Coulomb systems: a problem at vanishing particle densities*, J. Statist. Phys., 41 (1985), pp. 975–1013.

[19] O. Lanford and D. Robinson, *Mean entropy of states in quantum-statistical mechanics.*, J. Math. Phys., 9 (1968), pp. 1120–1125.

[20] E. H. Lieb, *The stability of matter*, Rev. Mod. Phys., 48 (1976), pp. 553–569.

[21] ———, *The stability of matter: from atoms to stars*, Bull. Amer. Math. Soc. (N.S.), 22 (1990), pp. 1–49.

[22] ———, *The stability of matter and quantum electrodynamics.* Buschhorn, Gerd W. (ed.) et al., Fundamental physics – Heisenberg and beyond. Papers from the Werner Heisenberg centennial symposium *development in modern physics*, Munich, Germany, December 5, 2001. Berlin: Springer. 53-68, 2004. `arXiv:math-ph/0401004`.

[23] E. H. Lieb and J. L. Lebowitz, *The constitution of matter: Existence of thermodynamics for systems composed of electrons and nuclei*, Advances in Math., 9 (1972), pp. 316–398.

[24] E. H. Lieb and M. B. Ruskai, *A fundamental property of quantum-mechanical entropy*, Phys. Rev. Lett., 30 (1973), pp. 434–436.

[25] ———, *Proof of the strong subadditivity of quantum-mechanical entropy*, J. Math. Phys., 14 (1973), pp. 1938–1941. With an appendix by B. Simon.

[26] E. H. Lieb and B. Simon, *The Thomas-Fermi theory of atoms, molecules and solids*, Advances in Math., 23 (1977), pp. 22–116.

[27] E. H. Lieb and J. P. Solovej, *Ground state energy of the two-component charged Bose gas.*, Commun. Math. Phys., 252 (2004), pp. 485–534.

[28] E. H. Lieb and W. E. Thirring, *Bound on kinetic energy of fermions which proves stability of matter*, Phys. Rev. Lett., 35 (1975), pp. 687–689.

[29] E. H. Lieb and H.-T. Yau, *The stability and instability of relativistic matter*, Commun. Math. Phys., 118 (1988), pp. 177–213.

[30] M. Loss, *Stability of matter.* Mourão, José C. (ed.) et al., Prospects in mathematical physics. Young researchers symposium of the 14th international congress on mathematical physics, Lisbon, Portugal, July 25–26, 2003. Providence, RI: American Mathematical Society (AMS). Contemporary Mathematics 435, 171-198, 2007.

[31] D. Robinson and D. Ruelle, *Mean entropy of states in classical statistical mechanics*, Commun. Math. Phys., 5 (1967), pp. 288–300.

[32] D. Ruelle, *Statistical mechanics. Rigorous results*, Singapore: World Scientific. London: Imperial College Press , 1999.

[33] J. P. Solovej, *The energy of charged matter.* Zambrini, Jean-Claude (ed.), XIVth international congress on mathematical physics (ICMP 2003), Lisbon, Portugal, 28 July – 2 August 2003. Selected papers based on the presentation at the conference. Hackensack, NJ: World Scientific. 113-129 (2005)., 2005.

[34] ———, *Upper bounds to the ground state energies of the one- and two-component charged Bose gases*, Commun. Math. Phys., 266 (2006), pp. 797–818.

[35] A. Wehrl, *General properties of entropy*, Rev. Modern Phys., 50 (1978), pp. 221–260.

4

No-Photon Quantum Electrodynamics in the Hartree-Fock Approximation

For heavy atoms, it is necessary to take relativistic effects into account. However there is no equivalent of the well-known N-body (non-relativistic) Schrödinger theory involving the Dirac operator, because of its negative spectrum. The correct theory is Quantum Electrodynamics (QED). This theory has a remarkable predictive power but its description in terms of perturbation theory restricts its range of applicability. In fact a mathematically consistent formulation of the nonperturbative theory is still unknown. On the other hand, effective models deduced from nonrelativistic theories (like the Dirac-Hartree-Fock model [84, 28]) suffer from inconsistencies: for instance a ground state never minimizes the physical energy which is always unbounded from below.

Here we study a *variational* model based on a physical energy which can be minimized to obtain the ground state in a chosen charge sector. Our model describes the behavior of a finite number of particles (electrons), coupled to that of the Dirac sea which can become polarized. Our results are fully non-perturbative.

The chapter is organized as follows: in Section 4.1 we explain (formally) how the Hartree-Fock approximation can be derived from first-principle QED. The Hartree-Fock free vacuum is constructed in Section 4.2 and the Bogoliubov-Dirac-Fock (BDF) model which describes any Hartree-Fock state by using the free vacuum as a reference state is introduced in Section 4.3. Existence of minimizers are studied with or without charge constraints in Sections 4.4 and 4.5. In Section 4.6 we neglect the exchange term and get some more properties of minimizers, in particular linked with charge renormalization. The time-dependent equation associated with the model is studied in Section 4.7. Lastly, the BDF model at positive temperature is described in Section 4.8.

This chapter is an enhanced version of [45] and of the last part of [27].

4.1 The Hartree-Fock approximation of QED

4.1.1 The Dirac operator

Dirac derived his operator in 1928 [17], starting from the usual classical expression of the energy of a free relativistic particle of momentum $p \in \mathbb{R}^3$ and mass m

$$E^2 = c^2|p|^2 + m^2c^4 \qquad (4.1)$$

(c is the speed of light), and imposing the necessary relativistic invariances. By means of the usual identification

$$p \longleftrightarrow -i\hbar\nabla$$

where $\hbar$ is Planck's constant, he found that an adequate observable for describing the energy of the free particle should therefore be a self-adjoint operator D_c satisfying the equation

$$(D_c)^2 = -c^2\hbar^2\Delta + m^2c^4. \tag{4.2}$$

Taking the locality principle into account, Dirac proposed to look for a local operator which is first order with respect to $p = -i\hbar\nabla$:

$$D_c \ = -ic\hbar\, \boldsymbol{\alpha}\cdot\nabla + mc^2\beta = \ -ic\hbar\sum_{k=1}^{3}\alpha_k\partial_k \ + \ mc^2\beta, \tag{4.3}$$

where α_1, α_2, α_3 and β are hermitian matrices which have to satisfy the following anticommutation relations:

$$\left\{\begin{array}{rcl} \alpha_k\alpha_\ell + \alpha_\ell\alpha_k & = & 2\,\delta_{k\ell}\,\mathbb{1}, \\ \alpha_k\beta + \beta\alpha_k & = & 0, \\ \beta^2 & = & \mathbb{1}. \end{array}\right. \tag{4.4}$$

It can be proved [87] that the smallest dimension in which (4.4) can take place is 4 (i.e. α_1, α_2, α_3 and β should be 4×4 hermitian matrices), meaning that D_c has to act on $L^2(\mathbb{R}^3, \mathbb{C}^4)$. The usual representation in 2×2 blocks is given by

$$\beta = \begin{pmatrix} I_2 & 0 \\ 0 & -I_2 \end{pmatrix}, \quad \alpha_k = \begin{pmatrix} 0 & \sigma_k \\ \sigma_k & 0 \end{pmatrix} \qquad (k = 1, 2, 3)\,, \tag{4.5}$$

where the Pauli matrices are defined as

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{4.6}$$

The main unusual feature of the Dirac operator is the fact that its spectrum is not bounded from below:

$$\sigma(D_c) = (-\infty, -mc^2] \cup [mc^2, \infty). \tag{4.7}$$

Compared with non-relativistic theories in which the Schrödinger operator $-\Delta/(2m)$ appears instead of D_c, property (4.7) leads to important physical, mathematical and numerical difficulties. Indeed, if one simply replaces $-\Delta/(2m)$ by D_c in the energies or operators which are commonly used in the non-relativistic case, one obtains energies which are not bounded from below.

Although there is no observable electron of negative energy, the negative spectrum plays an important role in physics. Dirac himself suspected that the negative spectrum of his operator could generate new interesting physical phenomena, and he proposed in 1930 the following interpretation [18, 19, 20]:

> "We make the assumption that, in the world as we know it, nearly all the states of negative energy for the electrons are occupied, with just one electron in each state, and that a uniform filling of all the negative-energy states is completely unobservable to us." [20]

Physically, one therefore has to imagine that the vacuum (called the *Dirac sea*) is filled with infinitely many virtual particles occupying the negative energy states. With this conjecture, a real free electron cannot be in a negative state due to the Pauli principle which forbids it to be in the same state as a virtual electron of the Dirac sea.

It was realized just after Dirac's discovery that, for consistency of the theory, the vacuum should not be considered as a totally virtual physical object which does not interact with the real

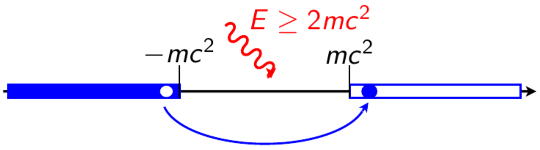


Figure 4.1: Creation of an electron-positron pair.

particles. Dirac himself [18, 20, 19] conjectured the existence of surprising physical effects as a consequence of his theory, which were then experimentally confirmed. First, the virtual electrons of the Dirac sea can feel an external field and they will react to this field accordingly, i.e. the vacuum will become *polarized*. This polarization is then felt by the real particles and one therefore is led to consider a coupled system 'Dirac sea + real particles'. From the experimental viewpoint, vacuum polarization plays a rather small role for the calculation of the Lamb shift of hydrogen but it is important for high-Z atoms [68] and it is even a crucial physical effect for muonic atoms [32, 36]. Second, in the presence of strong external fields, the vacuum could react so importantly that an electron-positron pair can be spontaneously created [69, 38, 75, 76].

The mathematical difficulties of a model aiming at describing both the Dirac sea and the real particles are important, for one has to deal at the same time with infinitely many particles (the real ones and the virtual ones of the Dirac sea). In the following, we present a Hartree-Fock (mean-field) type model for this problem, which has been mathematically studied in collaborations with Hainzl, Séré, Solovej, Gravejat, Seiringer and Sparber [42, 43, 46, 44, 47, 41, 37]. The results have already been reviewed in [45] and in the last part of [27]. The model under consideration is inspired of an important physical article by Chaix and Iracane [15] in which the possibility that a bounded-below energy could be obtained by adding vacuum polarization was first proposed. But the equations of this so-called Bogoliubov-Dirac-Fock model were already known in QED [75].

For the sake of simplicity, we take $c = 1$ except when explicitly mentioned. In this case an additional parameter $\alpha = e^2$ appears, where e is the (bare) charge of the electron.

4.1.2 The Hamiltonian of QED

In the next two sections, we formally derive the Hartree-Fock energy of no-photon QED, see Formula (4.12). It is a functional (which can only be made rigorous in finite volume) whose main variable $\gamma = P - 1/2$ is a self-adjoint operator satisfying $-1/2 \leq \gamma \leq 1/2$. Here $0 \leq P \leq 1$ is the usual one-body density matrix of the Hartree-Fock state and the subtraction of $1/2$ is a kind of renormalization whose usefulness will be explained below. The reader not familiar with second-quantization can at first jump directly to the remarks after Formula (4.12) in which we explain our methodology.

We start with the formal QED Hamiltonian written in Coulomb gauge, in the presence of an external electromagnetic potential (V, a), see [48, 49, 80, 78, 5]

$$\mathbb{H}^{V,a} = \int \Psi^*(x) \left[\boldsymbol{\alpha} \cdot (-i\nabla - \mathbb{A}(x) - a(x)) + m\beta\right] \Psi(x)\, dx + \int V(x)\rho(x)\, dx \\ + \frac{\alpha}{2} \iint \frac{\rho(x)\rho(y)}{|x-y|} dx\, dy + H_f. \quad (4.8)$$

In this formula, $\Psi(x)$ is the second quantized field operator which annihilates an electron at x and satisfies the anticommutation relation

$$\Psi^*(x)_\sigma \Psi(y)_\nu + \Psi(y)_\nu \Psi^*(x)_\sigma = 2\delta_{\sigma,\nu}\delta(x-y). \quad (4.9)$$

We use here physicists' notation ($\Psi(x)_\sigma$ is an operator-valued distribution). It is possible to give a real mathematical meaning to these formulas, assuming the system is confined to a box with an ultraviolet cut-off [46]. We will however detail this later on and continue to write formal expressions for the moment.

The operator $\rho(x)$ is the *density operator* defined by

$$\rho(x) = \sum_{\sigma=1}^{4} \frac{[\Psi^*_\sigma(x), \Psi_\sigma(x)]}{2} \tag{4.10}$$

where $[a,b] = ab - ba$. The operator H_f describes the kinetic energy of the photons:

$$H_f = \frac{1}{8\pi\alpha}\int \left(|\nabla \times \mathbb{A}(x)|^2 + |\mathbb{E}_t(x)|^2\right)\,dx = \frac{1}{\alpha}\sum_{\lambda=1,2}\int_{\mathbb{R}^3} dk\,|k|a^*_\lambda(k)a_\lambda(k) + \mathrm{Cte}$$

(Cte indicates a constant which diverges in infinite volume). The operators $\mathbb{A}(x)$ and $\mathbb{E}_t(x)$ are the electromagnetic field operators for the photons and $a^*_\lambda(k)$ is the creation operator of a photon with momentum k and polarization λ. In (4.8), (V,a) is an external electromagnetic potential. The potential V can for instance be created by a set of nuclei, a physical situation that we will consider in the rest of the chapter. The Hamiltonian $\mathbb{H}^{V,a}$ formally acts on the Fock space,

$$\mathcal{F} = \mathcal{F}_{\mathrm{e}} \otimes \mathcal{F}_{\mathrm{ph}}$$

where $\mathcal{F}_{\mathrm{e}}$ is the fermionic Fock space for the electrons and $\mathcal{F}_{\mathrm{ph}}$ is the bosonic Fock space for the photons.

We emphasize that (4.8) does not contain any normal-ordering or notion of (bare) electrons and positrons: $\Psi(x)$ can annihilate electrons of negative kinetic energy. The distinction between electrons and positrons should be a result of the theory and not an input. The commutator used in the formula (4.10) of $\rho(x)$ is a kind of renormalization, independent of any reference. It is due to Heisenberg [48] (see also [70, Eq. (96)]) and it is necessary for a covariant formulation of QED, see [78, Eq. (1.14)] and [21, Eq. (38)]. More precisely, the Hamiltonian $\mathbb{H}^{V,a}$ possesses the interesting property of being invariant under charge conjugation since the following relations hold formally

$$\mathscr{C}\rho(x)\mathscr{C}^{-1} = -\rho(x), \qquad \mathscr{C}\mathbb{H}^{V,a}\mathscr{C}^{-1} = \mathbb{H}^{-V,a},$$

where $\mathscr{C}$ is the charge conjugation operator acting on the Fock space. Associated with $\rho(x)$ we define an average number operator (counted relatively to $I/2$)

$$\mathcal{N} := \int \rho(x)\,dx$$

which satisfies $\mathscr{C}\mathcal{N}\mathscr{C}^{-1} = -\mathcal{N}$.

In Quantum Electrodynamics, the *vacuum* is defined to be the *state of lowest energy* of the system under consideration. If $V = 0$ and $a = 0$, one obtains the *free vacuum*. It is expected that the free vacuum Ω_0 is never charged, $\langle\Omega_0|\mathcal{N}|\Omega_0\rangle = 0$. It is also expected that the free vacuum is invariant by translation, in a sense to be made more precise later. If the external fields do not vanish, the state of lowest energy Ω is the *polarized vacuum* in the presence of V and a. It might be charged when the fields are strong enough. For instance in the presence of a very strong external field V created by positively charged particles, it is expected that one (or even several) electron-positron pair(s) will be created [76]. However the positron(s) will escape to infinity whereas the electron(s) will stay in a neighborhood of the positive density of charge inducing the electrostatic field. At the end we will have $\langle\Omega|\mathcal{N}|\Omega\rangle = N > 0$, the number of such created pairs.[1]

---

[1] Note that this phenomenon is different from the usual formulation of pair-creation based on the time-dependent equation [69, 72, 85, 86]. In our case we may talk about "static pair creation" as opposed to "dynamic pair creation". In the static case the ground state does not depend continuously on the strength of the external field: it suddenly jumps from one charge sector to another one provided the associated energy has become lower. For a system without interaction, static pair creation will occur when the first eigenvalue equals 0. In the dynamic case, it was proved that pair creation occurs when the first eigenvalue dives into the continuum [69, 72] i.e. when it passes -1 (however this only makes sense in an adiabatic limit). The positron is then able to radiate through the continuum. The difference between the two concepts was already mentioned earlier in [77]. The study of dynamic pair-creation within one of the nonlinear models presented in this chapter is a very interesting challenge.

If one wants to describe atomic systems (like electrons in a molecule), one has to find the state of lowest energy *in a fixed charge sector*, i.e. under the constraint that $\langle\Omega|\mathcal{N}|\Omega\rangle = N$ where $\mathcal{N}$ is defined as above. Intuitively if the external fields are not too strong, this should lead to a system of N electrons (or $-N$ positrons when $N < 0$), coupled to the Dirac sea.

4.1.3 Derivation of the Hartree-Fock energy

In our study of the QED Hamiltonian $\mathbb{H}^{V,a}$, we shall make two approximations:

- we neglect photons and assume there is no external magnetic field, $a \equiv 0$;
- we work in a mean-field theory, i.e. we restrict the Hamiltonian to Hartree-Fock states.

These approximations are of a different importance. Neglecting photons is of course a very rough approximation as it will forbid us to describe important physical effects occurring in QED like the self-energies of the electrons, the biggest contribution to the Lamb shift. But we do that only for mathematical reasons: we were not yet able to extend most of the results presented below when photons are taken into account. *Formally*, a large part of our study is exactly the same with photons (when they are treated by a mean-field procedure). We hope to come back to this point in the near future.

The second approximation which we make by restricting ourselves to Hartree-Fock states is more fundamental and many of our results are specific to this case. Nevertheless, some of our general ideas may be applicable to the full QED model.

Let us recall that the electronic *one-body density matrix* (two point function) of any electronic state $|\Omega\rangle \in \mathcal{F}_{\rm e}$ is defined as

$$P(x,y)_{\sigma,\sigma'} = \langle\Omega|\Psi^*(x)_\sigma\Psi(y)_{\sigma'}|\Omega\rangle$$

and it satisfies $0 \leq P \leq I$ where I is the identity operator. In view of (4.10), it is natural to introduce a *renormalized one-body density matrix*

$$\gamma(x,y)_{\sigma,\sigma'} = \left\langle\Omega\left|\frac{[\Psi(x)^*_\sigma, \Psi(y)_{\sigma'}]}{2}\right|\Omega\right\rangle.$$

By (4.9), we obtain the simple relation

$$\gamma = P - \frac{I}{2}$$

hence $-I/2 \leq \gamma \leq I/2$. Electronic Hartree-Fock states form a subset $\{|\Omega_P\rangle\} \subset \mathcal{F}_{\rm e}$ of states which are completely determined by their density matrix P (or equivalently by their renormalized density matrix $\gamma = P - I/2$). Recall that if

$$|\Omega\rangle = \varphi_1 \wedge \cdots \wedge \varphi_N$$

is a Hartree-Fock states with N occupied orbitals $\varphi_1, ..., \varphi_N$, then the associated density matrix P is just the orthogonal projector on $\mathrm{Span}(\varphi_1, ..., \varphi_N)$:

$$P = \sum_{i=1}^N |\varphi_i\rangle\langle\varphi_i|.$$

For a *formal* Hartree-Fock state with infinitely many occupied orbitals

$$|\Omega\rangle = \varphi_1 \wedge \cdots \wedge \varphi_N \wedge \cdots$$

we also obtain

$$P = \sum_{\rm occ} |\varphi_i\rangle\langle\varphi_i|.$$

Hence

$$\gamma = P - \frac{I}{2} = \frac{P - P^{\perp}}{2} = \frac{1}{2}\left(\sum_{\text{occ}} |\varphi_i\rangle\langle\varphi_i| - \sum_{\text{unocc}} |\varphi_i\rangle\langle\varphi_i|\right).$$

The associated density of charge is formally given by

$$\rho_\gamma(x) = \langle\Omega|\rho(x)|\Omega\rangle = \frac{1}{2}\left(\sum_{\text{occ}} |\varphi_i(x)|^2 - \sum_{\text{unocc}} |\varphi_i(x)|^2\right). \tag{4.11}$$

Now we can (formally) compute the energy of any state $|\Omega_P\rangle \otimes |0\rangle$ where $|\Omega_P\rangle$ is a Hartree-Fock state in $\mathcal{F}_{\text{e}}$ and $|0\rangle \in \mathcal{F}_{\text{ph}}$ is the photonic vacuum. We obtain

$$\boxed{\langle 0| \otimes \langle\Omega_P|\mathbb{H}^{V,0}|\Omega_P\rangle \otimes |0\rangle = \mathcal{E}^{\nu}_{\text{HF}}(P - I/2) + \text{Cte}}$$

where Cte is a constant (diverging in the infinite volume limit) and

$$\mathcal{E}^{\nu}_{\text{HF}}(\gamma) = \text{Tr}(D^0\gamma) - \alpha \iint \frac{\rho_\gamma(x)\nu(y)}{|x-y|}\,dx\,dy + \frac{\alpha}{2}\iint \frac{\rho_\gamma(x)\rho_\gamma(y)}{|x-y|}dx\,dy - \frac{\alpha}{2}\iint \frac{|\gamma(x,y)|^2}{|x-y|}dx\,dy. \tag{4.12}$$

Here and in the rest of the whole chapter we will consider for convenience an external electrostatic field of the form

$$V = -\alpha\nu * \frac{1}{|x|}$$

which is the usual Coulomb field induced by a distribution of charge ν.

The reader can recognize in (4.12) the well-known Hartree-Fock energy introduced in Chapter 1, but applied to the renormalized density matrix $\gamma = P - I/2$ instead of the usual density matrix P. The last two terms of the first line are respectively the kinetic energy and the interaction energy of the electrons with the external potential induced by the charge distribution ν. In the second line appear respectively the so-called *direct* and *exchange* terms. In Relativistic Density Functional Theory [24, 25], the exchange term is approximated by a function of ρ_γ and its derivatives only.

Note that we have restricted ourselves to quasi-free states having a vanishing *pairing density matrix*. Let us recall from Chapter 2 that the latter is defined for any state Ω as [4]:

$$p(x,y)_{\sigma,\sigma'} = \langle\Omega|\Psi^*(x)_\sigma\Psi^*(y)_{\sigma'}|\Omega\rangle.$$

The energy of a general quasi-free state is then equal to (4.12) plus $(\alpha/2)\iint |p(x,y)|^2|x-y|^{-1}dx\,dy$. Hence one always decreases the energy by removing the pairing density matrix (as in usual generalized Hartree-Fock theory with a repulsive potential [4]).

Any stationary point of the above energy formally satisfies the first order equation (written in terms of the usual density matrix $P = \gamma + I/2$)

$$\left[P, F_{P-I/2}\right] = 0$$

where $F_{P-I/2}$ is the Fock operator

$$F_{P-I/2} = D^0 + \alpha\left(\rho_{[P-I/2]} - \nu\right) * \frac{1}{|x|} - \alpha\frac{(P-I/2)(x,y)}{|x-y|}.$$

For a minimizer (in a chosen charge sector), one will have the more precise equation

$$P = \chi_{(-\infty,\mu]}\left(F_{P-I/2}\right) \tag{4.13}$$

where μ is a Fermi level which is adjusted to get the correct charge (it vanishes if there is no charge constraint). Saying differently, one obtains a Hartree-Fock state with infinitely many occupied orbitals, all having an energy $\leq \mu$. We shall give a precise interpretation of this equation later on.

It is time to worry about the mathematical meaning of the formulas we have formally derived up to now, in particular the definition of the energy (4.12). Unfortunately, the latter does not make any sense for the following reason: when P is an orthogonal projector (as this is usually the case for HF minimizers), $\gamma = P - I/2$ is never a compact operator in an infinite dimension space. Hence none of the terms appearing in (4.12) has a clear mathematical meaning. Formally, one has $\mathcal{E}^\nu_{\rm HF}(P - I/2) = -\infty$ for any density matrix P.

In [46], we proposed to overcome this difficulty in the following way: we restrict the whole system to a box of size L with periodic boundary conditions and an ultraviolet cut-off Λ in the Fourier domain. Then all the above formulas make perfectly sense because we are in a finite-dimensional setting. In particular one can define minimizers of the HF energy with or without the external field V, with or without a charge constraint. Then, we look at the limit of the minimizer in the considered class when the size of the box grows, $L \to \infty$, but the cut-off Λ stays fixed. The limit (if it exists) is the formal minimizer of the unbounded below energy $\mathcal{E}^\nu_{\rm HF}$.

Notice the ultraviolet cut-off Λ is fixed during the whole study. It is only at the very end that we can tackle the difficult task to remove it by *renormalization*.

We explain all that in details in the next sections.

4.2 Definition of the free vacuum

We start by explaining how the free vacuum was constructed in [46], in the case $\nu = 0$. This is done by defining the energy (4.12) for operators P acting on the finite-dimensional space

$$\begin{aligned}\mathfrak{H}^L_\Lambda &:= \left\{f \in L^2_{\rm per}([-L/2, L/2)^3, \mathbb{C}^4),\ \operatorname{supp}(\widehat{f}) \subset B(0,\Lambda)\right\}\\ &= \operatorname{span}\left\{\exp(ik\cdot x),\ k \in (2\pi/L)\mathbb{Z}^3 \cap B(0,\Lambda)\right\}.\end{aligned}$$

To define the energy properly, it is necessary to periodize the Coulomb potential as follows:

$$W_L(x) = \frac{1}{L^3}\left(\sum_{\substack{k\in(2\pi)\mathbb{Z}^3/L\\ k\neq 0}} \frac{4\pi}{|k|^2}e^{ik\cdot x} + wL^2\right),$$

where w is some constant which is chosen such that $\min_{\mathcal{C}_L} W_L = 0$ for any L. The Dirac operator D^0 is also easily defined on $\mathfrak{H}^L_\Lambda$: it is just the multiplication of the Fourier coefficients by $(D^0(k))_{k\in 2\pi\mathbb{Z}^3/L}$. Then, one introduces

$$\begin{aligned}\mathcal{E}^0_L(\Gamma) := \operatorname{Tr}_{\mathfrak{H}^L_\Lambda}(D^0\Gamma) + +\frac{\alpha}{2}\iint_{\mathcal{C}_L\times\mathcal{C}_L} W_L(x-y)\rho_\Gamma(x)\rho_\Gamma(y)\,dx\,dy&\\ - \frac{\alpha}{2}\iint_{\mathcal{C}_L\times\mathcal{C}_L} W_L(x-y)|\Gamma(x,y)|^2 dx\,dy\,,&\quad (4.14)\end{aligned}$$

for any self-adjoint operator Γ acting on $\mathfrak{H}^L_\Lambda$. The kernel $\Gamma(x,y)$ of Γ is easily defined in the Fourier basis since $\mathfrak{H}^L_\Lambda$ is finite-dimensional. Its density ρ_Γ is then defined as $\rho_\Gamma(x) := \operatorname{Tr}_{\mathbb{C}^4}(\Gamma(x,x))$. A translation-invariant operator T acting on $\mathfrak{H}^L_\Lambda$ is by definition a multiplication operator in the Fourier domain. In this case, one has $T(x,y) = f(x-y)$ for some f and therefore ρ_T is constant. The identity of $\mathfrak{H}^L_\Lambda$, denoted by I^L_Λ is an example of a translation-invariant operator.

It is possible to define the QED Hamiltonian without photons in the box in the same way, see [46]. Notice the fermionic Fock space built on the one-body space $\mathfrak{H}^L_\Lambda$ is also finite-dimensional.

The first result proved in [46] is the following:

Theorem 4.1. (QED mean-field minimizer in a box [46]) *Assume that* $0 \leq \alpha < 4/\pi$, $\Lambda > 0$ *and that* L *is large enough. Then the functional* $\mathcal{E}_L^0$ *has a unique minimizer* Γ_L^0 *on the convex set*

$$\mathcal{G}_\Lambda^L := \left\{\Gamma \in \mathcal{L}(\mathfrak{H}_\Lambda^L),\ \Gamma^* = \Gamma,\ -I_\Lambda^L/2 \leq \Gamma \leq I_\Lambda^L/2\right\}.$$

It is invariant by translation and satisfies $\rho_{\Gamma_L^0} \equiv 0$. *Moreover, it takes the form* $\Gamma_L^0 = \mathcal{P}_L^0 - I_\Lambda^L/2$ *where* $\mathcal{P}_L^0$ *is an orthogonal projector on* $\mathfrak{H}_\Lambda^L$.

Of course, it can easily be shown that $\mathcal{P}_L^0$ satisfies an equation similar to (4.13) with V removed and $1/|x|$ replaced by W_L. But we do not give more details since we are more interested in the limit of $\mathcal{P}_L^0$ as $L \to \infty$. We note that although the HF energy (in a box) is not convex, we have been able to prove that there is no symmetry breaking in the sense that its minimizer is invariant by translation and unique. We are not aware of any previous result of this kind for Hartree-Fock-type theories including the non-convex exchange term.[2]

In order to state the thermodynamic limit correctly, we first need to introduce the translation-invariant projector $\mathcal{P}_-^0$ acting on $\mathfrak{H}_\Lambda$, which will be the limit of the sequence $(\mathcal{P}_L^0)_L$. The identity of $\mathfrak{H}_\Lambda$ is denoted by I_Λ. We introduce

$$\boxed{\mathcal{T}(A) = \frac{1}{(2\pi)^3}\int_{B(0,\Lambda)} \mathrm{Tr}_{\mathbb{C}^4}[D^0(p)A(p)]dp - \frac{\alpha}{(2\pi)^5}\iint_{B(0,\Lambda)^2} \frac{\mathrm{Tr}_{\mathbb{C}^4}[A(p)A(q)]}{|p-q|^2}dp\,dq} \tag{4.15}$$

for any A belonging to the convex set

$$\mathcal{A}_\Lambda := \{A \text{ translation-invariant on } \mathfrak{H}_\Lambda,\ A^* = A,\ -I_\Lambda/2 \leq A \leq I_\Lambda/2\}.$$

It will be shown in Theorem 4.3 below that $\mathcal{T}$ represents the energy per unit volume of translation-invariant operators. For this reason, one now considers the minimization of $\mathcal{T}$ on $\mathcal{A}_\Lambda$. The following was proved in [46]:

Theorem 4.2. (Definition of the free vacuum [46]) *Assume that* $0 \leq \alpha < 4/\pi$ *and* $\Lambda > 0$. *Then* $\mathcal{T}$ *possesses a unique global minimizer* Γ^0 *on* $\mathcal{A}_\Lambda$. *It satisfies the self-consistent equation*

$$\begin{cases} \Gamma^0 = -\dfrac{\mathrm{sgn}(\mathcal{D}^0)}{2}, \\ \mathcal{D}^0 = D^0 - \alpha\dfrac{\Gamma^0(x,y)}{|x-y|} \end{cases} \tag{4.16}$$

or, written in terms of the translation-invariant projector $\mathcal{P}_-^0 = \Gamma^0 + I_\Lambda/2$,

$$\mathcal{P}_-^0 = \chi_{(-\infty,0)}\left(\mathcal{D}^0\right). \tag{4.17}$$

Moreover, $\mathcal{D}^0$ *takes the special form, in the Fourier domain,*

$$\mathcal{D}^0(p) = g_1(|p|)\boldsymbol{\alpha}\cdot p + g_0(|p|)\beta \tag{4.18}$$

where g_0, $g_1 \in L^\infty([0,\Lambda),\mathbb{R})$ *are such that* $1 \leq g_1(x) \leq g_0(x)$ *for any* $x \in [0,\Lambda)$, *and therefore*

$$|D^0(p)|^2 \leq |\mathcal{D}^0(p)|^2 \leq g_0(|p|)|D^0(p)|^2. \tag{4.19}$$

The self-consistent equation (4.16) has already been solved by Lieb and Siedentop in a different context [62]. They used a fixed point method only valid when $\alpha\log\Lambda \leq C$ for some constant C.

As shown by the next result, the negative spectral projector $\mathcal{P}_-^0$ of the Dirac-type operator $\mathcal{D}^0$ represents the *free vacuum*, as it is the limit of the sequence $\mathcal{P}_L^0$ when $L \to \infty$. An important

---

[2] The main idea was indeed taken from [2]: it consists in controlling the exchange term by the kinetic energy.

property of Γ^0 showing the usefulness of the subtraction of half the identity is the following. Due to

$$\mathcal{P}^0_-(p) - I_\Lambda(p)/2 = \Gamma^0(p) = -\frac{g_1(|p|)}{2\sqrt{g_1(|p|)^2|p|^2 + g_0(|p|)^2}}\boldsymbol{\alpha}\cdot p - \frac{g_0(|p|)}{2\sqrt{g_1(|p|)^2|p|^2 + g_0(|p|)^2}}\beta, \quad (4.20)$$

one infers

$$\mathrm{Tr}_{\mathbb{C}^4}(\Gamma^0(p)) = \mathrm{Tr}_{\mathbb{C}^4}[(\mathcal{P}^0_- - I_\Lambda/2)(p)] = 0\,,$$

for any $p \in B(0,\Lambda)$, the Dirac matrices being trace-less. This has the important consequence that the (constant) density of charge of the free vacuum vanishes:

$$\rho_{\Gamma^0} \equiv (2\pi)^{-3}\int_{B(0,\Lambda)} \mathrm{Tr}_{\mathbb{C}^4}(\Gamma^0(p))\,dp = 0.$$

This formally means that

$$\text{“}\,\langle\Omega_0|\mathcal{N}|\Omega_0\rangle = \int_{\mathbb{R}^3}\rho_{\Gamma^0}\,dx = 0\,\text{”} \quad (4.21)$$

where Ω_0 is the (formal) Hartree-Fock vacuum state in the Fock space, as desired.

In QED, the Feynman propagator at equal times

$$S_F(x,y;t_x = t_y) := i\gamma(x,y)\beta$$

is often expressed using the Källén-Lehmann representation [53, 59, 5], based on relativistic invariances. Although our model is not fully relativistically invariant (we discard photons and use an ultraviolet cut-off Λ) and is only defined in the mean-field approximation, our solution (4.20) has exactly the form which may be derived from the Källén-Lehmann representation for the equal time propagator. In four-dimensional full QED, a self-consistent equation similar to (4.16) is well-known and used. These so-called Schwinger-Dyson equations [79, 22] have been approximately solved for the free vacuum case first by Landau *et al.* in [57, 58], and then by many authors (see, e.g., [52, 40, 1]).

We notice that $\mathcal{P}^0_-$ is *not* Dirac's original choice P^0_- (except when $\alpha = 0$) because the interaction between the particles (the virtual and the real ones) is taken into account by the model. Notice also that Equation (4.17) is exactly the same as (4.13) with $V = 0$ and $\mu = 0$, due to (4.16).

As a consequence of (4.18), the spectrum of $\mathcal{D}^0$ is

$$\sigma(\mathcal{D}^0) = \left\{\pm\sqrt{g_0(|p|)^2 + g_1(|p|)^2|p|^2},\ p \in B(0,\Lambda)\right\}.$$

It has a gap which is greater than the one of D^0, by (4.19):

$$1 \le m(\alpha) := \min\sigma(|\mathcal{D}^0|). \quad (4.22)$$

In [44], it is proved that when $\alpha \ll 1$, then $m(\alpha) = g_0(0)$ and conjectured this is true for any $0 \le \alpha < 4/\pi$. Notice that the following expansion is known [62, 46]: $g_0(0) = 1 + \frac{\alpha}{\pi}\mathrm{arcsinh}(\Lambda) + \mathrm{O}(\alpha^2)$.

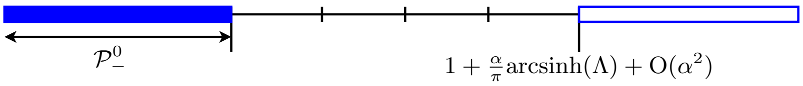


Figure 4.2: The free vacuum $\mathcal{P}^0_-$ is the negative spectral projector of the translation-invariant operator $\mathcal{D}^0$.

We are now able to state the thermodynamic limit, as proved in [46]:

Theorem 4.3. (Thermodynamic limit in the free case [46]) *Assume that* $0 \leq \alpha < 4/\pi$ *and* $\Lambda > 0$. *Then, one has*

$$\lim_{L\to\infty} \frac{\mathcal{E}^0_L(\Gamma^0_L)}{L^3} = \min_{\mathcal{A}_\Lambda} \mathcal{T}\,,$$

where we recall that Γ^0_L *is the unique minimizer of* $\mathcal{E}^0_L$ *defined in Theorem 4.1. Moreover,* $\mathcal{P}^0_L = \Gamma^0_L + I^L_\Lambda/2$ *converges to* $\mathcal{P}^0_-$ *in the following sense:*

$$\lim_{L\to\infty} \left\|\mathcal{P}^0_L - \mathcal{P}^0_-\right\|_{\mathfrak{S}_\infty(\mathfrak{H}^L_\Lambda)} = \lim_{L\to\infty} \sup_{p\in(2\pi\mathbb{Z}^3/L)\cap B(0,\Lambda)} |\mathcal{P}^0_L(p) - \mathcal{P}^0_-(p)| = 0.$$

4.3 The Bogoliubov-Dirac-Fock model

Now that the free vacuum $\mathcal{P}^0_-$ has been correctly defined, we will be able to introduce the Bogoliubov-Dirac-Fock (BDF) energy as studied in [42, 43, 44, 46]. Formally, it measures the energy (4.12) of a state P, relatively to the (infinite) energy of the free vacuum $\mathcal{P}^0_-$. It depends on $P - \mathcal{P}^0_-$ and reads formally:

$$\begin{aligned} \text{“}\mathcal{E}^\nu(P - \mathcal{P}^0_-) &= \mathcal{E}^\nu_{\rm HF}(P - I_\Lambda/2) - \mathcal{E}^0_{\rm HF}(\mathcal{P}^0_- - I_\Lambda/2) \qquad (4.23)\\ &= \operatorname{Tr}\left(\mathcal{D}^0 Q\right) - \alpha \iint_{\mathbb{R}^3\times\mathbb{R}^3} \frac{\rho_Q(x)\nu(y)}{|x-y|}dx\,dy \\ &\quad + \frac{\alpha}{2}\iint_{\mathbb{R}^3\times\mathbb{R}^3} \frac{\rho_Q(x)\rho_Q(y)}{|x-y|}dx\,dy - \frac{\alpha}{2}\iint_{\mathbb{R}^3\times\mathbb{R}^3} \frac{|Q(x,y)|^2}{|x-y|}dx\,dy\text{”}, \qquad (4.24) \end{aligned}$$

with $Q = P - \mathcal{P}^0_-$. This new energy looks again like a Hartree-Fock type functional except that our main variable is $Q = P - \mathcal{P}^0_-$, which measures the difference between our state and the (physically unobservable) translation-invariant free vacuum $\mathcal{P}^0_-$. Note also that the translation-invariant mean-field operator $\mathcal{D}^0$ appears[3].

As we will see later on, the above formal computation can again be justified by a thermodynamic limit. We will proceed as follows: first we explain how we have been able to give a proper mathematical meaning to the above energy and its minimizers. Then we show that the so-constructed states are indeed the limits of minimizing states in the thermodynamic limit $L \to \infty$.

The energy (4.24) was introduced and studied by Chaix-Iracane in [15] (see also Chaix-Iracane-Lions [16]). An adequate mathematical formalism was then provided by Bach, Barbaroux, Helffer and Siedentop [2] in the free case $\nu = 0$, and by Hainzl, Lewin and Séré [42, 43] in the external field case $\nu \neq 0$. However, in all these works a simplified version was considered: $\mathcal{D}^0$ and $\mathcal{P}^0_-$ were replaced by Dirac's choice D^0 and P^0_-. As mentioned in [46], although the choice of $\mathcal{P}^0_-$ for the free vacuum is better physically, the two models are essentially the same from the mathematical point of view: the main results of [2, 42, 43] were easily generalized in [46] to treat the model in which the better $\mathcal{D}^0$ and $\mathcal{P}^0_-$ are used.

What is gained with (4.24) is that Q can now be a compact operator (it will be Hilbert-Schmidt, indeed) and, thanks to the Fourier cut-off Λ, many of the terms in (4.24) will be mathematically well-defined. However, there is still a notable difficulty in defining the BDF energy correctly, i.e. finding what is the correct functional setting. First one might be tempted to restrict the functional to operators Q which are trace-class [74, 81], so that the density ρ_Q and the kinetic energy term $\operatorname{Tr}(\mathcal{D}^0 Q)$ are both well-defined. However we will see below that minimizers of the energy are *never trace-class*. This can be seen formally by expanding the self-consistent equation on first order in α. This peculiar mathematical property is not purely technical: as we will explain below, it is the main explanation of *charge renormalization*, an important physical issue in QED.

For this reason the formal BDF energy has to be defined on a larger set than only trace-class operators. We refer to Appendix A where the correct functional analysis setting is presented in

---

[3] It is obtained from the cross terms which are obtained when expanding the formal difference $\mathcal{E}^\nu_{\rm HF}(\mathcal{P}^0_- - I_\Lambda/2 + Q) - \mathcal{E}^0_{\rm HF}(\mathcal{P}^0_- - I_\Lambda/2)$ in terms of Q.

details. The main idea is to generalize the trace functional in order to define correctly the kinetic energy $\mathrm{Tr}\left(\mathcal{D}^0 Q\right)$. This is done by introducing the following space

$$\mathfrak{S}_1^{\mathcal{P}^0_-}(\mathfrak{H}_\Lambda) := \left\{Q \in \mathfrak{S}_2(\mathfrak{H}_\Lambda) \mid \mathcal{P}^0_- Q \mathcal{P}^0_-,\ \mathcal{P}^0_+ Q \mathcal{P}^0_+ \in \mathfrak{S}_1(\mathfrak{H}_\Lambda)\right\},$$

with the the usual notation

$$\mathfrak{S}_p(\mathfrak{H}_\Lambda) := \{A \in \mathcal{L}(\mathfrak{H}_\Lambda),\ \mathrm{Tr}(|A|^p) < \infty\}.$$

An operator Q belonging to $\mathfrak{S}_1^{\mathcal{P}^0_-}(\mathfrak{H}_\Lambda)$ is said to be *$\mathcal{P}^0_-$-trace class.* For any such $Q \in \mathfrak{S}_1^{\mathcal{P}^0_-}(\mathfrak{H}_\Lambda)$, we then define its *$\mathcal{P}^0_-$-trace* as

$$\mathrm{Tr}_{\mathcal{P}^0_-}(Q) := \mathrm{Tr}(\mathcal{P}^0_- Q \mathcal{P}^0_-) + \mathrm{Tr}(\mathcal{P}^0_+ Q \mathcal{P}^0_+).$$

Due to the fact that the free vacuum has a vanishing charge (4.21), $\mathrm{Tr}_{\mathcal{P}^0_-}(Q)$ can be interpreted as the charge of our state $P = Q + \mathcal{P}^0_-$. Our generalization of the trace functional has many interesting features which are proved in [42] and reviewed in Appendix A.

Definition of the density for states in $\mathfrak{S}_1^{\mathcal{P}^0_-}(\mathfrak{H}_\Lambda)$

Thanks to the cut-off in Fourier space, the charge density ρ_Q of an operator $Q \in \mathfrak{S}_1^{\mathcal{P}^0_-}(\mathfrak{H}_\Lambda)$ is well-defined in $L^2(\mathbb{R}^3, \mathbb{R})$, via

$$\widehat{\rho_Q}(k) = (2\pi)^{-3/2} \int_{\substack{|p+k/2|\le\Lambda \\ |p-k/2|\le\Lambda}} \mathrm{Tr}_{\mathbb{C}^4}\left(\widehat{Q}(p+k/2, p-k/2)\right) dp.$$

Let us introduce the following notation:

$$D(f,g) = 4\pi \int \frac{\overline{\widehat{f}(k)}\widehat{g}(k)}{|k|^2} dk\,,$$

for any $(f,g) \in L^2(\mathbb{R}^3, \mathbb{R})^2$, which coincides with $\iint_{\mathbb{R}^3\times\mathbb{R}^3} f(x)g(y)|x-y|^{-1}dx\,dy$ when f and g are smooth enough. We also define the so-called *Coulomb space*

$$\mathcal{C} = \{f \mid D(f,f) < \infty\}\,, \tag{4.25}$$

which is the natural space for defining the terms depending on ρ_Q in the energy. The following Lemma was proved in [41]:

Lemma 4.1 (The density is in $\mathcal{C} \cap L^2$ [41, Lemma 1])**.** *Let $0 \le \alpha < 4/\pi$ and $\Lambda > 0$. The map $Q \in \mathfrak{S}_1^{\mathcal{P}^0_-}(\mathfrak{H}_\Lambda) \mapsto \rho_Q \in \mathcal{C} \cap L^2$ is continuous: there exists a constant $C = C(\alpha, \Lambda)$ such that*

$$\|\rho_Q\|_{\mathcal{C}} + \|\rho_Q\|_{L^2} \le C(\Lambda)\, \|Q\|_{\mathfrak{S}_1^{\mathcal{P}^0_-}(\mathfrak{H}_\Lambda)}.$$

Definition of the BDF energy

It is now possible to define the Bogoliubov-Dirac-Fock energy as in [42, 43, 46]

$$\boxed{\mathcal{E}^\nu(Q) := \mathrm{Tr}_{\mathcal{P}^0_-}(\mathcal{D}^0 Q) - \alpha D(\rho_Q, \nu) + \frac{\alpha}{2} D(\rho_Q, \rho_Q) - \frac{\alpha}{2} \iint_{\mathbb{R}^6} \frac{|Q(x,y)|^2}{|x-y|} dx\,dy\,,} \tag{4.26}$$

where

$$Q \in \mathcal{K} := \left\{Q \in \mathfrak{S}_1^{\mathcal{P}^0_-}(\mathfrak{H}_\Lambda),\ -\mathcal{P}^0_- \le Q \le \mathcal{P}^0_+\right\}. \tag{4.27}$$

As we have stated in Lemma 4.1, we know that $\rho_Q \in \mathcal{C}$ when $Q \in \mathfrak{S}_1^{\mathcal{P}^0_-}(\mathfrak{H}_\Lambda)$, hence $\mathcal{E}^\nu$ is well defined when $\nu \in \mathcal{C}$. The following was proved:

Theorem 4.4 (The BDF energy is bounded-below [16, 2, 42, 43, 46])**.** *Assume that* $0 \leq \alpha < 4/\pi$, $\Lambda > 0$ *and that* $\nu \in \mathcal{C}$.

(*i*) *One has*

$$\forall Q \in \mathcal{K}, \qquad \mathcal{E}^\nu(Q) + \frac{\alpha}{2} D(\nu,\nu) \geq 0\,, \tag{4.28}$$

and therefore $\mathcal{E}^\nu$ *is bounded from below on* $\mathcal{K}$.

(*ii*) *If moreover* $\nu = 0$, *then* $\mathcal{E}^0$ *is non-negative on* $\mathcal{K}$, 0 *being its unique minimizer.*

The boundedness from below of the BDF energy is an essential feature of the theory. It shows the usefulness of the inclusion of the vacuum effects in the model. The interpretation of (*ii*) is the following: by (4.23), it proves that the free vacuum $\mathcal{P}^0_-$ is the unique minimizer of the (formal) QED energy in the set of all the projectors P which are such that $P - \mathcal{P}^0_- \in \mathcal{K}$. In the previous subsection (Theorem 4.2), it was also proved that $\mathcal{P}^0_-$ is the unique minimizer of the energy per unit volume $\mathcal{T}$. These are two different ways of giving a mathematical meaning to the fact that $\mathcal{P}^0_-$ is the unique minimizer of the QED energy when no external field is present.

The case $\nu = 0$, (*ii*) in Theorem 4.4, was proved by Bach, Barbaroux, Helffer and Siedentop [2]. In this paper, the authors also study a relativistic model for $\nu \neq 0$, but with vacuum polarization neglected (in the spirit of a paper of Mittlemann [67], see [27]). They were inspired of a paper by Chaix, Iracane and Lions [16]. Then, it has been argued in [42, 43, 46] that the proof of the case $\nu \neq 0$, (*i*) in Theorem 4.4, is a trivial adaptation of [2].

Now that $\mathcal{E}^\nu$ has been shown to be bounded-below, it is natural to try to minimize it. Actually, we shall be interested in two minimization problems. The first is the global minimization of $\mathcal{E}^\nu$ in the whole set $\mathcal{K}$. As mentioned above, a global minimizer of the mean-field QED energy $\mathcal{E}^\nu_{\rm HF}$ and therefore of the BDF energy $\mathcal{E}^\nu$ (they formally differ by an infinite constant !) is interpreted as the *polarized vacuum* in the external electrostatic field $-\alpha\nu * |\cdot|^{-1}$. If one wants to describe a system of charge $-eN$, one has to minimize $\mathcal{E}^\nu$ in the Nth charge sector:

$$\mathcal{K}(N) := \left\{ Q \in \mathcal{K} \mid \mathrm{Tr}_{\mathcal{P}^0_-}(Q) = N \right\}.$$

These two minimization problems will be tackled in the following two sections.

Let us now give the proof of Theorem 4.4, which is a simple adaptation of arguments of [2].

Proof of Theorem 4.4. Let $Q \in \mathcal{K}$, and therefore satisfying the operator inequality

$$-\mathcal{P}^0_- \leq Q \leq \mathcal{P}^0_+. \tag{4.29}$$

It is easily proved that (4.29) is equivalent to

$$Q^2 \leq Q^{++} - Q^{--} \tag{4.30}$$

where $Q^{++} := \mathcal{P}^0_+ Q \mathcal{P}^0_+$ and $Q^{--} := \mathcal{P}^0_- Q \mathcal{P}^0_-$. This now implies that

$$0 \leq \mathrm{Tr}(|\mathcal{D}^0|Q^2) \leq \mathrm{Tr}_{\mathcal{P}^0_-}(\mathcal{D}^0 Q) \tag{4.31}$$

which shows that the kinetic energy is non-negative. We now use Kato's inequality (2.2) and Equation (4.19) to obtain

$$\iint_{\mathbb{R}^3\times\mathbb{R}^3} \frac{|Q(x,y)|^2}{|x-y|} dx\,dy \leq \frac{\pi}{2}\,\mathrm{Tr}(|D^0|Q^2) \leq \frac{\pi}{2}\,\mathrm{Tr}(|\mathcal{D}^0|Q^2).$$

Together with (4.31), this implies

$$\mathcal{E}^\nu(Q) \geq \left(1 - \alpha\frac{\pi}{4}\right) \mathrm{Tr}(|\mathcal{D}^0|Q^2) - \frac{\alpha}{2} D(\nu,\nu),$$

which easily ends the proof of Theorem 4.4. □

4.4 Global minimization of $\mathcal{E}^\nu$: the polarized vacuum

The existence of a global minimizer of $\mathcal{E}^\nu$ has been proved with Hainzl and Séré, first in [42] by a fixed-point argument valid only when $\alpha\sqrt{\log\Lambda} \leq C_1$ and $\alpha D(\nu,\nu)^{1/2} \leq C_2$, and then by a global minimization procedure in [43], valid for any cut-off Λ and $0 \leq \alpha < 4/\pi$. The precise statement of the latter is the following:

Theorem 4.5 (Definition of the polarized vacuum [42, 43, 46])**.** *Assume that* $0 \leq \alpha < 4/\pi$, $\Lambda > 0$ *and that* $\nu \in \mathcal{C}$. *Then* $\mathcal{E}^\nu$ *possesses a minimizer* $Q_{\rm vac}$ *on* $\mathcal{K}$ *such that* $P_{\rm vac} = Q_{\rm vac} + \mathcal{P}^0_-$ *is an orthogonal projector satisfying the self-consistent equation*

$$P_{\rm vac} = \chi_{(-\infty,0)}\left(\mathcal{D}_{Q_{\rm vac}}\right), \tag{4.32}$$

where

$$\begin{aligned}\mathcal{D}_{Q_{\rm vac}} &= \mathcal{D}^0 + \alpha\left(\rho_{Q_{\rm vac}} - \nu\right) * \frac{1}{|\cdot|} - \alpha\frac{Q_{\rm vac}(x,y)}{|x-y|} &(4.33)\\ &= D^0 + \alpha\left(\rho_{[\bar{\mathcal{P}}_- - 1/2]} - \nu\right) * \frac{1}{|\cdot|} - \alpha\frac{(\bar{\mathcal{P}}_- - 1/2)(x,y)}{|x-y|}. &(4.34)\end{aligned}$$

Additionally, if α *and* ν *satisfy*

$$0 \leq \alpha\frac{\pi}{4}\left\{1 - \alpha\left(\frac{\pi}{2}\sqrt{\frac{\alpha/2}{1-\alpha\pi/4}} + \pi^{1/6}2^{11/6}\right)D(\nu,\nu)^{1/2}\right\}^{-1} \leq 1, \tag{4.35}$$

then this global minimizer $Q_{\rm vac}$ *is unique and the associated polarized vacuum is neutral, i.e.* $Q_{\rm vac} \in \mathcal{K}(0)$*:*

$$\mathrm{Tr}_{\mathcal{P}^0_-}(Q_{\rm vac}) = \mathrm{Tr}_{\mathcal{P}^0_-}(P_{\rm vac} - \mathcal{P}^0_-) = 0. \tag{4.36}$$

The proof consists in showing that $\mathcal{E}^\nu$ is lower semi-continuous for the weak-$*$ topology of $\mathcal{K}$. For this purpose, one shows that, in the electron-positron field, any mass escaping to infinity takes away a positive energy. This is the so-called dichotomy case of the concentration-compactness principle [64]. To prove (4.36), one first uses that $\mathrm{Tr}_{\mathcal{P}^0_-}(P_{\rm vac} - \mathcal{P}^0_-)$ is always an integer as stated in Lemma A.2 of Appendix A and one applies a continuation argument.

Notice that the definition (4.16) of $\mathcal{D}^0$ has been used to obtain (4.34) from (4.33). Of course, equations (4.32) and (4.34) are exactly the one we wanted to solve in the beginning (4.13). For not too strong external densities ν, a neutral vacuum is necessarily obtained, as shown by (4.36). But in general, a charged polarized vacuum could be found.

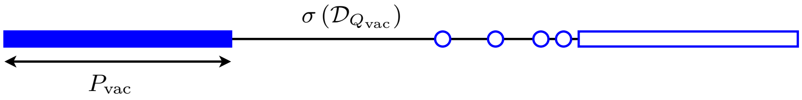


Figure 4.3: The polarized vacuum $P_{\rm vac}$ in the presence of the external density ν is the negative spectral projector of the mean-field operator $\mathcal{D}_{Q_{\rm vac}}$.

In [46], a thermodynamic limit was considered as for the free case $\nu = 0$, to justify the formal computation (4.24) when $\nu \neq 0$. As before, the QED energy is well-defined in a box $\mathcal{C}_L = [-L/2, L/2)^3$ with periodic boundary conditions and a cut-off in Fourier space by

$$\mathcal{E}^\nu_L(\Gamma) := \mathcal{E}^0_L(\Gamma) - \alpha\iint_{\mathcal{C}_L\times\mathcal{C}_L} W_L(x-y)\rho_\Gamma(x)\nu_L(y)\,dx\,dy, \tag{4.37}$$

where

$$\nu_L(x) = \frac{(2\pi)^{3/2}}{L^3}\sum_{k\in(2\pi\mathbb{Z}^3)/L}\widehat{\nu}(k)e^{ik\cdot x}, \tag{4.38}$$

(for simplicity, it is assumed that $\widehat{\nu}$ is a smooth function). Then the following was proved in [46].

Theorem 4.6 (Thermodynamic limit with external field [46])**.** *Assume that* $0 \leq \alpha < 4/\pi$*,* $\Lambda > 0$*,* $\nu \in \mathcal{C}$ *and that* $\widehat{\nu}$ *is continuous on* $B(0,\Lambda)$*. Then for any* L*,* $\mathcal{E}_L^\nu$ *possesses a minimizer* $\Gamma_L = \mathcal{P}_L - I_\Lambda^L/2$ *on* $\mathcal{G}_\Lambda^L$ *where* $\mathcal{P}_L$ *is an orthogonal projector, and one has*

$$\lim_{L\to\infty} \left\{\mathcal{E}_L^\nu(\Gamma_L) - \mathcal{E}_L^0(\Gamma_L^0)\right\} = \min\left\{\mathcal{E}^\nu(Q),\ Q \in \mathcal{K}\right\}. \tag{4.39}$$

Moreover, up to a subsequence, $Q_L(x,y) := (\Gamma_L - \Gamma_L^0)(x,y) = (\mathcal{P}_L - \mathcal{P}_L^0)(x,y)$ *converges uniformly on compact subsets of* $\mathbb{R}^6$ *to* $Q_{\rm vac}(x,y)$*, a minimizer of* $\mathcal{E}^\nu$ *on* $\mathcal{K}$*.*

4.5 Minimization of $\mathcal{E}^\nu$ in charge sectors

The previous subsection was devoted to the global minimization of the BDF energy. We now mention some results that have been obtained in [44] for the minimization with a charge constraint. It is believed that the charge constrained BDF model can be obtained as the thermodynamical limit of the full QED model in a fixed charge sector and posed in a box with periodic boundary conditions, but this has not been shown yet.

Due to the charge constraint and like for the Hartree-Fock model [63, 65, 60, 83] for instance, minimizers will not always exist for the BDF functional: it depends whether the external electrostatic potential created by the charge distribution ν is strong enough to be able to bind the N particles in the presence of the Dirac sea. On the other hand, it must not be too strong otherwise electron-positron pairs could appear.

We start with a general result proved in [44] providing the form of a minimizer, if it exists. To this end, we introduce the minimum energy in the Nth charge sector:

$$\boxed{E^\nu(N) := \inf\left\{\mathcal{E}^\nu(Q) \mid Q \in \mathcal{K}(N)\right\},} \tag{4.40}$$

where we recall that

$$\mathcal{K}(N) := \left\{Q \in \mathcal{K} \mid {\rm Tr}_{\mathcal{P}_-^0}(Q) = N\right\}.$$

In principle N could be any real number, but here, for simplicity, we shall restrict ourselves to integers, which corresponds to the physical case.

Theorem 4.7 (Self-Consistent Equation of a BDF Minimizer [44])**.** *Let* $0 \leq \alpha < 4/\pi$*,* $\Lambda > 0$*,* $\nu \in \mathcal{C}$ *and* $N \in \mathbb{Z}$*. Then any minimizer* Q *solution of the minimization problem* (4.40)*, if it exists, takes the form* $Q = P - \mathcal{P}_-^0$ *where*

$$P = \chi_{(-\infty,\mu]}(\mathcal{D}_Q) = \chi_{(-\infty,\mu]}\left(\mathcal{D}^0 + \alpha(\rho_Q - \nu) * 1/|\cdot| - \alpha\frac{Q(x,y)}{|x-y|}\right), \tag{4.41}$$

for some $\mu \in [-m(\alpha), m(\alpha)]$*.*

Recall that $m(\alpha)$ is the threshold of the free operator $\mathcal{D}^0$ defined in (4.22). We remark that (4.41) implicitly means that the last eigenvalue below μ of the mean-field operator $\mathcal{D}_Q$ is necessarily totally filled. This is a general fact for Hartree-Fock type theories [3]. For a minimizer of the form (4.41) and when $N, \mu > 0$, it is natural to consider the decomposition

$$P = \Pi + \chi_{[0\,,\,\mu]}(\mathcal{D}_Q),$$

where Π is the polarized Dirac sea:

$$\Pi := \chi_{(-\infty\,,\,0)}(\mathcal{D}_Q).$$

For not too strong external potentials, the vacuum will be neutral, *i.e.*

$${\rm Tr}_{\mathcal{P}_-^0}(\Pi - \mathcal{P}_-^0) = 0\,,$$

and therefore $\chi_{[0\,,\,\mu]}(\mathcal{D}_Q)$ will be a projector of rank N:

$$\chi_{[0,\mu]}(\mathcal{D}_Q) = \sum_{n=1}^{N} \varphi_n \otimes \varphi_n^*, \quad \mathcal{D}_Q\varphi_n = \varepsilon_n \varphi_n\,,$$

where $\varepsilon_1 \leq \cdots \leq \varepsilon_N$ are the first N positive eigenvalues of $\mathcal{D}_Q$ counted with their multiplicity. Notice that

$$\mathcal{D}_Q = D^0 + \alpha(\rho_\Phi - \nu) * \frac{1}{|\cdot|} - \alpha\frac{\gamma_\Phi(x,y)}{|x-y|} + \alpha\rho_{[\Pi - I_\Lambda/2]} * \frac{1}{|\cdot|} - \alpha\frac{(\Pi - I_\Lambda/2)(x,y)}{|x-y|}, \tag{4.42}$$

where

$$\gamma_\Phi := \chi_{[0,\mu]}(\mathcal{D}_Q) = \sum_{n=1}^{N} \varphi_n \otimes \varphi_n^*, \qquad \rho_\Phi(x) := \mathrm{Tr}_{\mathbb{C}^4}(\gamma_\Phi(x,x)) = \sum_{n=1}^{N} |\varphi_n(x)|^2.$$

In the terms of (4.42), the Dirac-Fock operator associated with $(\varphi_1, ..., \varphi_N)$ appears, see [28]. This shows that the electronic orbitals φ_i are solutions of a Dirac-Fock type equation in which the mean-field operator is perturbed by the (self-consistent) potentials of the Dirac sea $\Pi - I_\Lambda/2$. In practice, these potentials are small, and the DF equations are a good approximation of the BDF equations for the electronic orbitals. But the energy functionals behave in a completely different way: as explained in [28], the DF energy is strongly indefinite while the BDF energy is bounded below. The Dirac-Fock model is thus interpreted as a *non-variational* approximation of the mean-field model of no-photon QED [15].

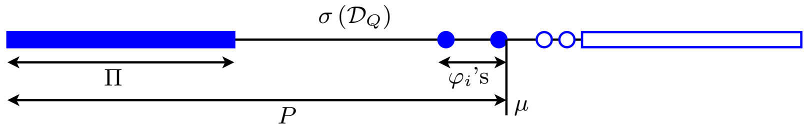


Figure 4.4: Decomposition of the system 'vacuum + N electrons' for the solution $P = \Pi + \gamma_\Phi$ in the Nth charge sector.

Concerning the existence of a minimizer, solution of (4.41), the following result was proved in [44]:

Theorem 4.8 (Binding Conditions and Existence of a BDF Minimizer [44])**.** *Let* $0 \leq \alpha < 4/\pi$*,* $\Lambda > 0$*,* $\nu \in \mathcal{C}$ *and* $N \in \mathbb{Z}$*. Then the following two assertions are equivalent:*

(H_1) $\qquad E^\nu(N) < \min\left\{E^\nu(N-K) + E^0(K),\ K \in \mathbb{Z} \setminus \{0\}\right\}.$

(H_2) *Each minimizing sequence* $(Q_n)_{n\geq 1}$ *for* $E^\nu(N)$ *is precompact in* $\mathcal{K}$ *and converges, up to a subsequence, to a minimizer* Q *of* $E^\nu(N)$*.*

Conditions like (H_1) appear classically when analyzing the compactness properties of minimizing sequences, for instance by using the concentration-compactness principle of P.-L. Lions [64]. They are also very classical for linear models in which the bottom of the essential spectrum has the form of the minimum in the right hand side of (H_1), as expressed by the HVZ Theorem [50, 89, 90]. Assume $N > 0$ for simplicity. When $0 < K \leq N$, (H_1) means that it is not favorable to let K electrons escape to infinity, while keeping $N - K$ electrons near the nuclei. When $K < 0$, it means that it is not favorable to let $|K|$ positrons escape to infinity, while keeping $N + |K|$ electrons near the nuclei. When $K > N$, it means that it is not favorable to let K electrons escape to infinity, while keeping $K - N$ positrons near the nuclei. When α is small enough and $N > 0$, it was shown in [44] that the separation of electron-positron pairs is not energetically favorable, so that one just needs to check (H_1) for $K = 1, 2, ..., N$.

To prove the existence of a minimizer, one can therefore prove that (H_1) holds. Two situations in which (H_1) is true have been provided in [44]. The first one is the case of weak coupling $\alpha \ll 1$ and $\alpha\nu = \bar{\nu}$ fixed (the charge N is also fixed). The following was proved:

Theorem 4.9 (Existence of a minimizer in the weak coupling limit [44]). *Assume that $\Lambda > 0$, that N is a non-negative integer, and that $\bar\nu \in \mathcal{C}$ is such that*

1. *the spectrum $\sigma(D^0 - \bar\nu * |\cdot|^{-1})$ contains at least N positive eigenvalues below 1,*
2. $\ker(D^0 - t\bar\nu * |\cdot|^{-1}) = \{0\}$ *for any $t \in [0,1]$.*

Then (H_1) holds in Theorem 4.8 for α small enough and $\alpha\nu = \bar\nu$, and therefore there exists a minimizer Q_α of $E^{\bar\nu/\alpha}(N)$. It takes the form

$$Q_\alpha = \chi_{(-\infty,0]}\left(\mathcal{D}_{Q_\alpha}\right) - \mathcal{P}^0_- + \chi_{(0,\mu_\alpha]}\left(\mathcal{D}_{Q_\alpha}\right) := Q_\alpha^{\rm vac} + \sum_{i=1}^N |\varphi_i^\alpha\rangle\langle\varphi_i^\alpha| \tag{4.43}$$

$$\mathcal{D}_{Q_\alpha}\varphi_i^\alpha = \varepsilon_i^\alpha\varphi_i^\alpha \tag{4.44}$$

where $\varepsilon_1^\alpha \leq \cdots \leq \varepsilon_N^\alpha$ are the N first positive eigenvalues of $\mathcal{D}_{Q_\alpha}$. Finally, for any sequence $\alpha_n \to 0$, $(\varphi_1^{\alpha_n}, ..., \varphi_N^{\alpha_n})$ converges (up to a subsequence) in $\mathfrak{H}_\Lambda$ to $(\varphi_1, ..., \varphi_N)$ which are N first eigenfunctions of $D^0 - \bar\nu * |\cdot|^{-1}$ and $Q_{\alpha_n}^{\rm vac}$ converges to $\chi_{(-\infty;0)}\left(D^0 - \bar\nu * |\cdot|^{-1}\right) - P^0_-$ in $\mathfrak{S}_2(\mathfrak{H}_\Lambda)$.

The second situation provided in [44] is the case of the non-relativistic regime $c \gg 1$. To state the result correctly, we reintroduce the speed of light c in the model (of course, we shall then take $\alpha = 1$). The expression of the energy and the definition of the free vacuum $\mathcal{P}^0_-$ (which of course then depends on c and the ultraviolet cut-off Λ) are straightforward. We denote by $E^\nu_{\alpha,c,\Lambda}(N)$ the minimum energy of the BDF functional depending on the parameters (α, c, Λ). The following was proved:

Theorem 4.10 (Existence of a minimizer in the non-relativistic limit [44]). *Assume that $\alpha = 1$ and that the ultraviolet cut-off is $\Lambda = \Lambda_0 c$ for some fixed Λ_0. Let $\nu \in \mathcal{C} \cap L^1(\mathbb{R}^3, \mathbb{R}^+)$ with $\int_{\mathbb{R}^3}\nu = Z$, and N a positive integer which is such that $Z > N - 1$. Then, for c large enough, (H_1) holds in Theorem 4.8 and therefore there exists a minimizer Q_c for $E^\nu_{1,c,\Lambda_0 c}(N)$. It takes the following form:*

$$Q_c = \chi_{(-\infty,0)}(\mathcal{D}_Q) - \mathcal{P}^0_- + \chi_{[0,\mu_c)}(\mathcal{D}_Q) = Q_c^{\rm vac} + \sum_{i=1}^N |\varphi_i^c\rangle\langle\varphi_i^c|\,,$$

and one has

$$\lim_{c\to\infty}\left\{E^\nu_{1,c,\Lambda_0 c}(N) - Ng_0(0)\right\} = \min_{\substack{0\leq\gamma\leq 1\\ \mathrm{Tr}(\gamma)=N}} \mathcal{E}_{\rm HF}(\gamma)$$

where $\mathcal{E}_{\rm HF}$ is the (spin-$1/2$) Hartree-Fock energy (1.8) *as defined in Chapter 1.*

Moreover, for any sequence $c_n \to \infty$, $(\varphi_1^{c_n}, ..., \varphi_N^{c_n})$ converges in $H^1(\mathbb{R}^3, \mathbb{C}^4)^N$ (up to a subsequence) towards $(\varphi_1, ..., \varphi_N)$ with $\varphi_i = \binom{\bar\varphi_i}{0}$, and where $\gamma = \sum_{i=1}^N |\bar\varphi_i\rangle\langle\bar\varphi_i|$ is a global minimizer of the Hartree-Fock energy.

We notice that this theorem is very similar to a result of [29] providing the convergence of the Dirac-Fock 'ground state' in the non-relativistic limit.

4.6 The reduced model and charge renormalization

4.6.1 Renormalization: generalities

In regular QED, the divergences of the (appropriately defined) physical measurable quantities are usually eliminated by means of a *mass* and a *charge renormalization.* The main idea is to assume that the parameters α and m appearing in the theory are indeed *bare parameters* which are not physically observable. The *physical parameters* are assumed to be functions of α, m and the cut-off Λ

$$\alpha_{\rm ph} = \alpha_{\rm ph}(\alpha, m, \Lambda), \qquad m_{\rm ph} = m_{\rm ph}(\alpha, m, \Lambda)$$

and equal the physical values obtained in experiment. These functions should be inverted in order to express the unknown bare quantities in term of the physical quantities

$$\alpha = \alpha(\alpha_{\rm ph}, m_{\rm ph}, \Lambda), \qquad m = m(\alpha_{\rm ph}, m_{\rm ph}, \Lambda). \tag{4.45}$$

Using these functions, one expects to remove (in some sense that needs to be precised) all divergences from physically measurable quantities.

Mass and charge renormalization however does not remove all divergences in the theory. Certain quantities, e.g. the bare Feynman propagator S_F (either at equal times or at general space time points), are still divergent. The expectation is that all these divergences cancel in physically measurable quantities and that they are therefore of no real relevance in formulating the theory.

Although there is no real need to do this, it is often convenient to introduce a renormalization of the bare Feynman propagator S_F. This is referred to as a *wavefunction renormalization*. In full QED [22] it is claimed that the divergence in the Feynman propagator may be removed by a *multiplicative renormalization* and that the renormalized propagator has the same pole near mass shell in 4-momentum space as a free propagator corresponding to a particle with the correct physical mass.

Note that in practice, this theoretical renormalization procedure is always used to justify the dropping of the divergent terms obtained at each order of the perturbation theory [22]. For this fact to be true, it is particularly important that renormalization can be expressed by means of *multiplicative* parameters in front of the different propagators [22].

In Hartree-Fock QED, it is not clear at all if the usual renormalization program of QED can be applied, especially when photons are not included. In [75, p. 194–195], it is argued that mass and charge renormalization is alone not enough to completely remove the divergences of the HF theory by means of multiplicative parameters.

In any case, the physical mass and charge have to be identified within the model. We propose the following definitions (one may expect that other physically reasonable definitions would lead to the same observable quantities in the limit $\Lambda \to \infty$). In relativistic quantum mechanics the physical mass is just the lowest energy of a free electron, hence

$$m_{\rm ph}(\alpha, m, \Lambda) := E^0(1) \tag{4.46}$$

which was defined in (4.40).

To define the physical coupling constant, we consider an extended nucleus of density ν, $\int \nu = Z$, and put it in the vacuum. Let $Q_{\rm vac} = P_{\rm vac} - \mathcal{P}^0_-$ be the polarized vacuum solution of (4.32). We assume that ν is not too strong such that the vacuum stays neutral, ${\rm Tr}_0(Q_{\rm vac}) = 0$. Of course in reality it is impossible to distinguish the nucleus from the vacuum and the charge which is observed far way from the nucleus is just

$$e\left(Z - \int_{\mathbb{R}^3} \rho_{Q_{\rm vac}}\right)$$

(provided $\rho_{Q_{\rm vac}}$ is an L^1 function). Hence we may define

$$\alpha_{\rm ph}(\alpha, m, \Lambda) := \alpha\left(1 - Z^{-1}\int_{\mathbb{R}^3} \rho_{Q_{\rm vac}}\right). \tag{4.47}$$

As we will see, the so-obtained formula is indeed independent of Z (if not, we would take its limit as $Z \to 0$).

It is very important to realize that charge renormalization is based on the fact that the operator $Q_{\rm vac}$ is *not* trace-class. If it were trace-class, one would of course have ${\rm Tr}_0(Q_{\rm vac}) = 0 = \int \rho_{Q_{\rm vac}}$, hence $\alpha_{\rm ph} = \alpha$. Therefore, the mathematical difficulty that a minimizer of the BDF energy is never trace-class (except when $\nu = 0$) is the origin of charge renormalization. Also this shows that in a finite dimensional space (for computational purpose for instance), renormalization is certainly more involved as all operators are trace class.

Both (4.46) and (4.47) would define $m_{\rm ph}$ and $\alpha_{\rm ph}$ as extremely complicated non-linear functions of α, m and Λ. A challenging task is to study the finiteness of measurable quantities like for instance the binding energy of an electron in the presence of an external field $E^\nu(1) - E^0(1)$, when $\alpha_{\rm ph}$ and $m_{\rm ph}$ are fixed to be the observed physical quantities. We do not know if this is possible when photons are not taken into account.

It is however possible to completely solve the above program for a (further) simplified theory called the *reduced Bogoliubov-Dirac-Fock*, as was done in [43, 37]. We explain that now.

4.6.2 Minimizers in the reduced BDF model and charge renormalization

Following [82] we consider a simplified model where the exchange term is neglected from the very beginning, i.e. in Formula (4.12). In Relativistic Density Functional Theory [24, 25], the exchange term is approximated by a function of the density only. The formal QED energy becomes:

$$\boxed{\mathcal{E}^\nu_{\rm rHF}(\gamma) = {\rm Tr}(D^0\gamma) - \alpha \iint \frac{\rho_\gamma(x)\nu(y)}{|x-y|}\,dx\,dy + \frac{\alpha}{2}\iint \frac{\rho_\gamma(x)\rho_\gamma(y)}{|x-y|}dx\,dy.} \tag{4.48}$$

The main advantage compared to the HF case is that this energy is now (formally) convex with respect to γ and strictly convex with respect to ρ_γ. This is a huge simplification which dramatically reduces the complexity of many proofs.

For this new functional one can follow the method of the previous paragraphs. One finds that the free vacuum is this time described by the usual Dirac sea, i.e. its density matrix is $P^0_- = \chi_{(-\infty,0)}(D^0)$. Hence, subtracting this energy to the energy of any state, one obtains the *reduced Bogoliubov-Dirac-Fock* energy:

$$\boxed{\mathcal{E}^\nu_{\rm r}(Q) := {\rm Tr}_{P^0_-}(D^0Q) - \alpha D(\rho_Q,\nu) + \frac{\alpha}{2}D(\rho_Q,\rho_Q),} \tag{4.49}$$

where similarly as before

$$Q \in \mathcal{K}_{\rm r} := \left\{Q \in \mathfrak{S}_1^{P^0_-}(\mathfrak{H}_\Lambda),\ -P^0_- \leq Q \leq P^0_+\right\}. \tag{4.50}$$

Again one can prove that for any $Q \in \mathcal{K}_{\rm r}$, one has automatically $\rho_Q \in \mathcal{C}$. Also we have the same lower bound

$$\forall Q \in \mathcal{K}_{\rm r}, \qquad \mathcal{E}^\nu_{\rm r}(Q) + \frac{\alpha}{2}D(\nu,\nu) \geq 0\,, \tag{4.51}$$

Since $\mathcal{E}^\nu_{\rm r}$ is convex on $\mathcal{K}_{\rm r}$ and strongly continuous, it is weakly lower semi-continuous, hence it has a global minimizer[4] $Q_{\rm vac}$, interpreted as the rHF polarized vacuum in the presence of the external field induced by the density ν. This was remarked in [43, Theorem 3]. Assuming that $\ker(D_{Q_{\rm vac}}) = \{0\}$ where

$$D_{Q_{\rm vac}} = D^0 + \alpha(\rho_{Q_{\rm vac}} - \nu) * |\cdot|^{-1}$$

is the mean field operator, then one can adapt the proof of Theorem 4.5 to get that $Q_{\rm vac}$ is unique and is a solution of the nonlinear equation $Q_{\rm vac} = \chi_{(-\infty,0]}(D_{Q_{\rm vac}}) - P^0_-$. The charge of the polarized vacuum is $-eq_0$ where

$$\boxed{q_0 = {\rm Tr}_{P^0_-}(Q_{\rm vac}).}$$

When $\alpha D(\nu,\nu)^{1/2}$ is not too large [43, Eq. (15)], it was proved that $q_0 = 0$. However in general electron-positron pairs can appear, giving rise to a charged vacuum. When $\ker(D_{Q_{\rm vac}}) \neq \{0\}$, then $\mathcal{E}^\nu_{\rm r}$ does not necessarily have a unique global minimizer on $\mathcal{K}_{\rm r}$, but it was proved in [37] that q_0 is anyway a uniquely defined quantity.

---

[4] We will use the same notation $Q_{\rm vac}$ as for the BDF model with exchange term, although, of course, the two states are different.

4.6.2.1 Existence of minimizers and their charge

Like in Section 4.5, we are interested in the following minimization problem

$$\boxed{E^\nu_{\rm r}(q)=\inf_{Q\in\mathcal{K}_{\rm r}(q)}\mathcal{E}^\nu_{\rm r}(Q)} \tag{4.52}$$

where the sector of charge $-eq$ is by definition

$$\mathcal{K}_{\rm r}(q):=\{Q\in\mathcal{K}_{\rm r},\ {\rm Tr}_{P^0_-}(Q)=q\}$$

and q is any real number. Of course in Physics $q\in\mathbb{Z}$ but it is convenient to allow any real value. It was proved in [37] that $q\to E^\nu_{\rm r}(q)$ is a Lipschitz and convex function. Notice that if $Q_{\rm vac}$ is a global minimizer of $\mathcal{E}^\nu_{\rm r}$ on $\mathcal{K}_{\rm r}$, then $q_0={\rm Tr}_{P^0_-}(Q_{\rm vac})$ minimizes $q\to E^\nu_{\rm r}(q)$.

The following theorem was proved in [37]:

Theorem 4.11 (Existence of atoms and molecules in the reduced BDF model [37])**.** *Let $\Lambda>0$, $\alpha\geq 0$, $\nu\in L^1(\mathbb{R}^3)\cap\mathcal{C}$ and denote $Z=\int_{\mathbb{R}^3}\nu\in\mathbb{R}$. Then there exists $q_m\in[-\infty,\infty)$ and $q_M\in[q_m,\infty]$ such that*

(i) $[q_m,q_M]$ is the largest interval on which $q\to E^\nu_{\rm r}(q)$ is strictly convex. If $q_M<\infty$, then $E^\nu_{\rm r}(q)=E^\nu_{\rm r}(q_M)+q-q_M$ for any $q>q_M$. If $q_M>-\infty$, then $E^\nu_{\rm r}(q)=E^\nu_{\rm r}(q_m)+q_m-q$ for any $q<q_m$;

(ii) the interval $[q_m,q_M]$ contains both Z and the unique minimizer q_0 of $q\to E^\nu_{\rm r}(q)$;

(iii) if $q\notin[q_m,q_M]$, then $\mathcal{E}^\nu_{\rm r}$ has no minimizer in the charge sector $\mathcal{K}_{\rm r}(q)$*;*

(iv) if $q\in[q_m,q_M]$, then $\mathcal{E}^\nu_{\rm r}$ has a minimizer Q in the charge sector $\mathcal{K}_{\rm r}(q)$*. This minimizer is not* a priori *unique but its associated density ρ_Q is uniquely determined. It is radially symmetric if ν is radially symmetric. The operator Q satisfies the self-consistent equation*

$$\begin{cases} Q+P^0_-=\chi_{(-\infty,\mu)}\left(D_Q\right)+\delta,\\ D_Q=D^0+\alpha(\rho_Q-\nu)*|\cdot|^{-1},\end{cases} \tag{4.53}$$

where $\mu\in[-1,1]$ is a Lagrange multiplier associated with the charge constraint and interpreted as a chemical potential, and δ satisfies $0\leq\delta\leq 1$ and ${\rm Ran}(\delta)\subseteq\ker(D_Q-\mu)$. If $\mu\in(-1,1)$, then δ has a finite rank. If $\mu\in\{-1,1\}$, then δ is trace-class.

Moreover, ρ_Q belongs to $L^1(\mathbb{R}^3)$ and satisfies

$$\boxed{\int_{\mathbb{R}^3}\rho_Q-Z=\frac{q-Z}{1+\alpha B_\Lambda(0)}} \tag{4.54}$$

where

$$B_\Lambda(0)=\frac{1}{\pi}\int_0^{\frac{\Lambda}{\sqrt{1+\Lambda^2}}}\frac{z^2-z^4/3}{1-z^2}dz=\frac{2}{3\pi}\log\Lambda-\frac{5}{9\pi}+\frac{2\log 2}{3\pi}+O(1/\Lambda^2).$$

The constant $B_\Lambda(0)$ is the value at zero of some real function B_Λ which will be defined later in Equation (4.61).

Equation (4.54) has an important physical interpretation: it gives us the formula of the physical coupling constant in terms of the bare coupling constant, as explained in the previous section. We obtain from (4.54)

$$\boxed{\alpha_{\rm ph}=\frac{\alpha}{1+\alpha B_\Lambda(0)}.} \tag{4.55}$$

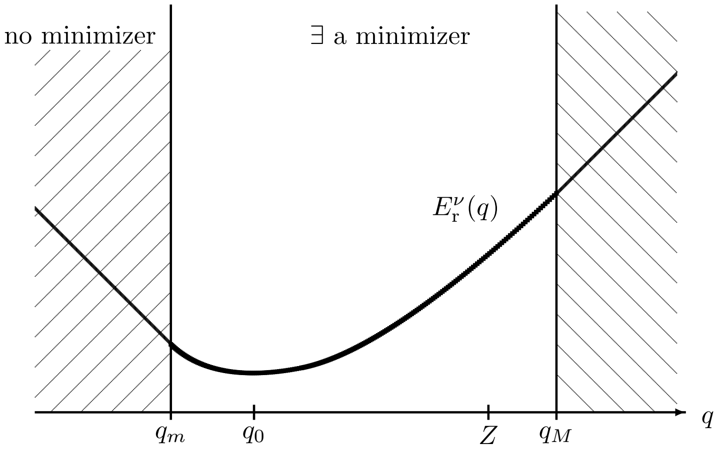


Figure 4.5: Schematic representation of the result for the rBDF model.

The above value of the observable α_ph is very well-known in QED, see e.g. [56, Eq. (8)] and [51, Eq. (7.18)]. It is important to note that the above formula is exact (within our model): contrarily to what is usually done in QED, it was obtained by a completely non perturbative method. Let us add that it was proved in [37] that

$$m_\mathrm{ph} = E_\mathrm{r}^\nu(1) = m$$

hence there is no mass renormalization in reduced BDF theory.

Equation (4.54) implies that a minimizer Q in the charge sector $q \neq Z$ is never trace-class, as this would imply $\mathrm{Tr}_{P^0_-} Q = \int_{\mathbb{R}^3} \rho_Q$ and contradict (4.54). This shows that the generalization of the reduced BDF energy $\mathcal{E}_\mathrm{r}^\nu$ to the Banach space $\mathfrak{S}_1^{P^0_-}(\mathfrak{H}_\Lambda)$ is mandatory, as no minimizer exists in the trace class. The mathematical difficulty that a minimizer is not trace-class is well interpreted physically in terms of charge renormalization.

When $q = Z$, it is in principle possible that a minimizer Q for $E_\mathrm{r}^\nu(q)$ is trace-class. This natural question has not been investigated.

We believe that a formula similar to (4.55) holds for the BDF model with exchange term (with a slightly modified constant $B_\Lambda(0)$ due to the replacement of D^0 by $\mathcal{D}^0$), but we have been unable to prove that $\rho_Q \in L^1$ in this case. Also we expect that there will be a mass renormalization with the exchange term. These important questions have not been solved yet.

4.6.2.2 An ionization estimate

It is well-known that in Quantum Electrodynamics a cut-off is mandatory [5, 51]. There are two sources of divergence in the Bogoliubov-Dirac-Fock model. The first is the negative continuous spectrum of the Dirac operator, which is cured by the subtraction of the (infinite) energy of the Dirac sea, as explained above. The second source of divergence is the rather slow growth of the Dirac operator for large momenta: D^0 only behaves linearly in p at infinity. This second issue is solved by the cut-off Λ in Fourier space,[5] which itself leads to renormalization.

There are many different ways to implement an ultraviolet cut-off. In the previous sections we have chosen above a simple "sharp" cut-off consisting in replacing the ambient Hilbert space $L^2(\mathbb{R}^3, \mathbb{C}^4)$ by the space of functions whose Fourier transform have their support in a ball of radius

---

[5] Note in Chapter 5 we will present a model for non-relativistic crystals which was studied in [9, 10]. It has a structure very similar to the BDF model of QED. But one main difference is that the second divergence does not occur, because of the presence of the Laplacian (instead of D^0) which has a faster growth at infinity. Hence there will not be any cut-off Chapter 5.

Λ. However, when looking at decay properties of the electronic density, it might be more adapted to impose a "smooth" cut-off (a similar remark was made by Lieb and Loss [61] in the context of non-relativistic QED). One possibility is to increase the growth of the Dirac operator at infinity, instead of changing the underlying Hilbert space. This has been proposed and studied in [37].

The idea is to replace D^0 by the following operator

$$D^0_\Lambda = (-i\boldsymbol{\alpha}\cdot\nabla+\beta)\left(1-\frac{\Delta}{\Lambda^2}\right).$$

The corresponding reduced BDF energy is then defined for states Q acting on the whole space $L^2(\mathbb{R}^3,\mathbb{C}^4)$ but with some growth conditions of the form

$$|D^0_\Lambda|^{1/2}Q\in\mathfrak{S}_2,\quad |D^0_\Lambda|^{1/2}Q^{++}|D^0_\Lambda|^{1/2}\in\mathfrak{S}_1,\quad |D^0_\Lambda|^{1/2}Q^{--}|D^0_\Lambda|^{1/2}\in\mathfrak{S}_1.$$

We refer to [37] for details. A result similar to Theorem 4.11 holds in this case, with a different constant $B'_\Lambda(0)=B_\Lambda(0)+o(1)$. The following was proved in [37]:

Theorem 4.12 (Estimates on q_m and q_M when $Z>0$, in the smooth cut-off case [37])**.** *We assume that D^0_Λ is defined as above. There exists universal constants $0<\theta_0<1$, $\alpha_0>0$ and $C>0$ such that the following holds. For any $0\le\alpha\le\alpha_0$, for any radial function $\nu\ge0$ in $L^1(\mathbb{R}^3)\cap\mathcal{C}$ such that $Z=\int\nu>0$ and $\alpha D(\nu,\nu)^{1/2}\le\theta_0<1$ and any cut-off $\Lambda\ge4$ such that $\alpha\log\Lambda<1/C$, the following estimate holds true:*

$$-C\frac{Z\alpha\log\Lambda+1/\Lambda+\alpha D(\nu,\nu)^{1/2}}{1-C\alpha\log\Lambda}\le q_m\le 0=q_0, \tag{4.56}$$

$$Z\le q_M\le\frac{2Z+C(Z\alpha\log\Lambda+1/\Lambda+\alpha D(\nu,\nu)^{1/2})}{1-C\alpha\log\Lambda}. \tag{4.57}$$

In a nonrelativistic limit in which one takes $\alpha\to0$, $\Lambda\to\infty$ such that $\alpha\log\Lambda\to0$ and ν fixed, one obtains the usual estimate of [60]

$$0=q_m=q_0<Z\le q_M\le 2Z.$$

4.6.3 Renormalization of the self-consistent equation

The renormalized charge is only observed far away from the nucleus. Close to it, one will observe a different behavior like the oscillations of the polarization of the vacuum $\rho_{Q_{\rm vac}}$. We now give some more details concerning the renormalization of charge and its interpretation in terms of the self-consistent equation. For simplicity, we consider the case of the polarized vacuum (the argument is the same in charge sectors).

Consider a small external density ν, $\int\nu=Z$ and let $Q_{\rm vac}$ be the associated polarized vacuum, with density $\rho_{\rm vac}:=\rho_{Q_{\rm vac}}$. As we have seen, the SCF equation satisfied by $Q_{\rm vac}$ reads

$$Q_{\rm vac}=\chi_{(-\infty;0]}(D_{Q_{\rm vac}})-\chi_{(-\infty;0]}(D^0). \tag{4.58}$$

When ν is small enough, we know that $0\notin\sigma(D_{Q_{\rm vac}})$ and that $Q_{\rm vac}$ is unique. We expand (4.58) in powers of α, using the resolvent representation [54, Section VI, Lemma 5.6] to derive the self-consistent equation for the density $\rho_{\rm vac}$

$$\rho_{\rm vac}(x)=-\frac{1}{2\pi}\int_{-\infty}^{\infty}d\eta\,{\rm Tr}\left[\frac{1}{D^0+\alpha(\rho_Q-\nu)*\frac{1}{|x|}+i\eta}-\frac{1}{D^0+i\eta}\right](x,x). \tag{4.59}$$

Applying the resolvent equation

$$\frac{1}{A-\alpha B}-\frac{1}{A}=\alpha\frac{1}{A}B\frac{1}{A}+\alpha^2\frac{1}{A}B\frac{1}{A}B\frac{1}{A}+\alpha^3\frac{1}{A}B\frac{1}{A}B\frac{1}{A}B\frac{1}{A-\alpha B}$$

and using Furry's Theorem [33], telling us that the corresponding α^2-term with two potentials vanish, we obtain

$$\rho_{\rm vac} = \alpha F_1[\rho_{\rm vac} - \nu] + F_3[\alpha(\rho_{\rm vac} - \nu)], \tag{4.60}$$

$$F_3[\rho](x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\eta \,\mathrm{Tr}\left[\frac{1}{D^0 + i\eta}\rho * \frac{1}{|x|}\frac{1}{D^0 + i\eta}\rho * \frac{1}{|x|}\frac{1}{D^0 + i\eta}\rho * \frac{1}{|x|}\frac{1}{D^0 + \rho * \frac{1}{|x|} + i\eta}\right](x,x).$$

As realized first by Dirac [20, 19] and Heisenberg [48], cf. also [34], the term $F_1[\rho]$ plays a particular role since it is logarithmically ultraviolet divergent. Following, e.g., Pauli-Rose [71], one evaluates in Fourier representation

$$\widehat{F}_1[\rho](k) = -\hat{\rho}(k)B_\Lambda(k),$$

where [71, Eq. (5)–(9)]

$$B^0_\Lambda(k) = \frac{-1}{\pi^2|k|^2}\int_{\substack{|\ell+k/2|\leq\Lambda,\\ |\ell-k/2|\leq\Lambda}} \frac{(\ell+k/2)\cdot(\ell-k/2) + 1 - E(\ell+k/2)E(\ell-k/2)}{E(\ell+k/2)E(\ell-k/2)(E(\ell+k/2)+E(\ell-k/2))}d\ell \tag{4.61}$$

and

$$B_\Lambda = B_\Lambda(0) = \frac{1}{\pi}\int_0^{\frac{\Lambda}{\sqrt{1+\Lambda^2}}} \frac{z^2 - z^4/3}{1 - z^2}\,dz = \frac{2}{3\pi}\log(\Lambda) - \frac{5}{9\pi} + \frac{2}{3\pi}\log 2 + O(1/\Lambda^2). \tag{4.62}$$

It can be seen that $B_\Lambda(k) = B_\Lambda - C_\Lambda(k)$, with

$$\lim_{\Lambda\to\infty} C_\Lambda(k) = C(k) = -\frac{1}{2\pi}\int_0^1 dx(1-x^2)\log[1 + k^2(1-x^2)/4], \tag{4.63}$$

which was first calculated by Serber and Uehling [80, 88].

We now show how to renormalize the SCF equation using (4.55). Denote $\rho = \rho_{\rm vac} - \nu$ the total (observable) density, then (4.60) can be rewritten in terms of ρ

$$\alpha\widehat{\rho} = -\alpha\widehat{\nu} - \alpha^2 B_\Lambda\widehat{\rho} + \alpha^2 C_\Lambda(k)\widehat{\rho} + \alpha\widehat{F}_3[\alpha\rho] \tag{4.64}$$

and

$$\alpha\hat{\rho} = -\frac{\alpha}{1+\alpha B_\Lambda}\widehat{\nu} + \frac{\alpha}{1+\alpha B_\Lambda}C_\Lambda(k)\alpha\widehat{\rho} + \frac{\alpha}{1+\alpha B_\Lambda}\widehat{F}_3[\alpha\rho]. \tag{4.65}$$

To perform our renormalization scheme we fix as physical (renormalized) objects $\alpha_{\rm ph}\rho_{\rm ph} = \alpha\rho$. Notice the renormalization of the density ρ is similar to a wavefunction renormalization of the (equal time) Feynman propagator as explained above. We can rewrite the self-consistent equation (4.64) as

$$\alpha_{\rm ph}\widehat{\rho}_{\rm ph} = -\alpha_{\rm ph}\widehat{\nu} + \alpha_{\rm ph}^2 C_\Lambda(k)\widehat{\rho}_{\rm ph} + \alpha_{\rm ph}\widehat{F}_3[\alpha_{\rm ph}\rho_{\rm ph}], \tag{4.66}$$

independently of the bare α. Notice that equation (4.66) satisfied by $\alpha_{\rm ph}\rho_{\rm ph}$ is exactly the same as equation (4.64) satisfied by $\alpha\rho$, but with the logarithmically divergent term $\alpha^2 B_\Lambda\widehat{\rho}$ dropped. Therefore, as usual in QED [22], the charge renormalization allows to simply justify the dropping of the divergent terms in the self-consistent equation. In practice [68], one would solve (4.66) perturbatively with $\alpha_{\rm ph} \simeq 1/137$ and with $C_\Lambda(k)$ replaced by its limit $C(k)$. However it is probably not expected that this scheme actually converges to a solution in the whole space [23].

Returning to the *effective* Hamiltonian $D^0 + \alpha(\rho_{\rm vac} - \nu) * 1/|x|$ and inserting (4.66), i.e. expressing in terms of the physical objects, we obtain

$$D^0 + \alpha_{\rm ph}\rho_{\rm ph} * \frac{1}{|x|} = D^0 - \alpha_{\rm ph}\nu * \frac{1}{|x|} + V_{\rm eff}, \tag{4.67}$$

with

$$V_{\rm eff} = \frac{2}{\pi^3}\mathcal{F}^{-1}\left[\frac{\alpha_{\rm ph}^2 C_\Lambda(k)\widehat{\rho}_{\rm ph}(k) + \alpha_r \hat{F}_3(\alpha_r\rho_{\rm ph})}{k^2}\right](x)$$

the effective self-consistent potential, where $\mathcal{F}^{-1}$ denotes the inverse Fourier transform. Notice, this equation is valid for any strength of the external potential. However, expanding $\rho_{\rm ph}$ in $\alpha_{\rm ph}$, we obtain to lowest order in $\alpha_{\rm ph}$

$$\begin{aligned} V_{\rm eff} &\simeq \alpha_{\rm ph}^2 \frac{2}{\pi^3}\mathcal{F}^{-1}\left[\frac{C_\Lambda(k)\widehat{\nu}(k)}{k^2}\right](x) \\ &\simeq \frac{\alpha_{\rm ph}^2}{3\pi}\int_1^\infty dt(t^2-1)^{1/2}\left[\frac{2}{t^2}+\frac{1}{t^4}\right]\int dx' e^{-2|x-x'|t}\frac{\nu(x')}{|x-x'|}, \end{aligned}$$

the Uehling potential [6].

The Landau pole

We notice that (4.55) can be written as

$$\alpha = \frac{\alpha_{\rm ph}}{1-\alpha_{\rm ph}B_\Lambda(0)}.$$

The fact that the denominator can go to zero is usually called the Landau pole. Also we see that

$$\alpha_{\rm ph}B_\Lambda(0) < 1 \tag{4.68}$$

which proves that $\alpha_{\rm ph} \to 0$ when $\Lambda \to \infty$, independently of α.

In physics, this "nullification" of the theory as the cut-off Λ diverges has been first suggested by Landau *et al.* [55, 56, 57, 58] and later studied by Pomeranchuk *et al.* [73]. We notice that with the usual value $\alpha_{\rm ph} \simeq \frac{1}{137}$, (4.68) leads to the physical bound $\Lambda < 10^{280}$ (in units of mc^2). This phenomenon has two consequences.

The first consequence is that when renormalizing observable quantities, it is not possible to fix $\alpha_{\rm ph}$ and take $\Lambda \to \infty$. Let $A(m,\alpha,\Lambda)$ be an observable quantity. Expressing m and α in terms of physical quantities we can write $A(m,\alpha,\Lambda) = \tilde{A}(m,\alpha_{\rm ph},\Lambda)$ (we have supposed that $m_{\rm ph} = m$). Unfortunately, it does not make sense to assume that $\tilde{A}(m,\alpha_{\rm ph},\Lambda)$ has a limit when $\alpha_{\rm ph}$ is fixed and $\Lambda \to \infty$: this function is only defined for Λ's which are such that $\alpha_{\rm ph}B_\Lambda(0) < 1$. However, let us assume that

$$\tilde{A}(m,\alpha_{\rm ph},\Lambda) = \sum_{k\geq 0} A_k(m,\Lambda)(\alpha_{\rm ph})^k.$$

By renormalization one sometimes means that each $A_k(m,\Lambda)$ should have a limit $A_k(m)$ when $\Lambda \to \infty$. However, it is probably not expected that the series $\sum_{k\geq 0} A_k(m)(\alpha_{\rm ph})^k$ will always have a positive radius of convergence [23].

The second consequence is a rather peculiar behavior of the minimization problem when $\Lambda \to \infty$ and α stays fixed. In [43, Thm 2], it was proved that for a fixed (and not too strong) external field $V = -\alpha\nu * \frac{1}{|x|}$ one has

$$\lim_{\Lambda\to\infty} \inf_{\substack{Q\in\mathfrak{S}_1^{P_-^0}(\mathfrak{H}_\Lambda) \\ -P_-^0\leq Q\leq P_+^0}} \mathcal{E}_{\rm r}^\nu(Q) = -\frac{1}{2}D(\nu,\nu),$$

hence the lower bound in (4.51) is optimal. This implies that the unique polarized vacuum P_Λ of the *reduced* BDF model satisfies

$$\lim_{\Lambda\to\infty} {\rm Tr}\left(P_\Lambda - P_-^0\right)^2 = 0 \quad \text{and} \quad \lim_{\Lambda\to\infty} D(\rho_{P_\Lambda - P_-^0} - \nu, \rho_{P_\Lambda - P_-^0} - \nu) = 0.$$

In words, when $\Lambda \to \infty$, the vacuum polarization density *totally cancels the external density* ν *in* $\mathcal{C}$, for $\rho_{P_\Lambda - P^0_-} \to \nu$. But since $P_\Lambda - P^0_- \to 0$, this means that in the limit $\Lambda \to \infty$, $P_\Lambda - P^0_-$ and its associated density become independent. Therefore, the minimization without cut-off makes no sense both from a mathematical and physical point of view. Indeed all this easily implies that when no cut-off is imposed and when $\nu \neq 0$, the infimum of the reduced BDF functional is *not attained.*

4.7 The time-dependent equation

The boundedness from below of the BDF energy (4.26) has been used in a work with Hainzl and Sparber [47] to prove the existence of *global-in-time* solutions to the time-dependent nonlinear equation associated with the BDF functional. The so-called *von Neumann equation* reads:

$$\begin{cases} i\dfrac{d}{dt}P(t) = [\mathcal{D}_{Q(t)}\,,\, P(t)], \\ \quad P(0) = P_I, \\ \quad Q(t) = P(t) - \mathcal{P}^0_- \in \mathcal{K}, \end{cases} \tag{4.69}$$

where $[\cdot,\cdot]$ is the usual commutator and

$$\mathcal{D}_Q := \mathcal{D}^0 + \alpha(\rho_Q - \nu) * \frac{1}{|\cdot|} - \alpha\frac{Q(x,y)}{|x-y|}$$

is the mean-field operator introduced before.[6] The following result was proved in [47]:

Theorem 4.13 (Existence of a unique global-in-time solution [47])**.** *Let* $\Lambda > 0$*,* $0 \le \alpha < 4/\pi$ *and* $\nu \in \mathcal{C}$*. Then, for any initial self-adjoint operator* P_I *such that* $Q_I = P_I - P^0 \in \mathcal{K}$*, there exists a unique maximal solution*

$$P(t) \in C^1\left([0,\infty), \mathcal{P}^0_- + \mathcal{K}\right)$$

of the Cauchy problem (4.69)*. Moreover, one has with* $Q(t) = P(t) - \mathcal{P}^0_-$

$$\mathrm{Tr}_{\mathcal{P}^0_-}(Q(t)) = \mathrm{Tr}_{\mathcal{P}^0_-}(Q_I) \quad \text{and} \quad \mathcal{E}^\nu(Q(t)) = \mathcal{E}^\nu(Q_I),$$

for all $t \in [0,\infty)$*.*

If the exchange term is neglected like in the original papers of Dirac [19, 20], the theorem is valid for *any* $\alpha \geq 0$.

The method of proof is similar to classical arguments already used for the Hartree-Fock theory, based on Schrödinger's equation, cf. [14, 13, 7, 8, 11] and the references given therein. We indeed solve the evolution-problem written in terms of $Q(t) = P(t) - \mathcal{P}^0_-$, i.e.

$$\begin{cases} i\dfrac{d}{dt}Q(t) = [\mathcal{D}_Q, Q] + [V_Q, \mathcal{P}^0_-], \\ \quad Q(0) = Q_I \in \mathfrak{S}_1^{\mathcal{P}^0_-}(\mathfrak{H}_\Lambda), \end{cases} \tag{4.70}$$

where

$$V_Q = \mathcal{P}_\Lambda\left(\alpha\big(\rho_Q - \nu\big) * \frac{1}{|\cdot|} - \alpha\frac{Q(x,y)}{|x-y|}\right)\mathcal{P}_\Lambda,$$

$\mathcal{P}_\Lambda$ being the projector onto $\mathfrak{H}_\Lambda$ in $L^2(\mathbb{R}^3, \mathbb{C}^4)$. The existence of a unique local-in-time solution in $\mathfrak{S}_1^{\mathcal{P}^0_-}(\mathfrak{H}_\Lambda)$ is obtained by classical arguments. Then, it is shown that the BDF energy is constant along this solutions, which consequently implies that it is indeed global-in-time, since $\mathcal{E}^\nu$ is bounded-below and coercive.

---

[6] Indeed, like in [42, 43], the model for which $\mathcal{P}^0_-$ and $\mathcal{D}^0$ are replaced by P^0_- and D^0 was considered, but the proof holds similarly in the case of (4.69).

The existence of a bounded below energy-functional, conserved along the solution $Q(t)$ is indeed a huge advantage of the BDF-model in comparison to other nonlinear Dirac equations, see, e.g., [26, 35, 39, 66]. There global-in-time solutions are obtained with quite strong restrictions on the considered initial data, due to the non-existence of such an *a priori* energy-estimate. More precisely, sufficient smallness assumptions within appropriate Sobolev norms are required in [26, 35, 66]. Similar assumptions for initial data corresponding to sufficient small scattering states at $t = +\infty$ are used in the related study of [30, 31]. Here, the proof is highly simplified by the important property that the BDF energy $\mathcal{E}^\nu$, which is conserved along solutions of (4.69), is bounded-below and coercive.

As a simple consequence of Theorems 4.5[7], 4.8 and 4.13, and of the method of Cazenave and Lions [12], we obtain the *orbital stability of the set of minimizers*. Let us recall that a set $\mathcal{M} \subset \mathcal{K}'$ is said to be orbitally stable in $\mathcal{K}' \subseteq \mathcal{K}$ if for any $\epsilon > 0$, there exists $\eta > 0$ such that for all $Q_I \in \mathcal{K}'$ with $\mathrm{dist}(Q_I, \mathcal{M}) \leq \eta$, if $t \mapsto Q(t)$ is the solution of (4.70) with initial data Q_I, we have $\mathrm{dist}(Q(t), \mathcal{M}) \leq \epsilon$ for all $t \in \mathbb{R}$. Here $\mathrm{dist}(Q, \mathcal{M}) := \inf_{Q' \in \mathcal{M}} \|Q - Q'\|_{\mathfrak{S}_1^{\mathcal{P}^0_-}(\mathfrak{H}_\Lambda)}$.

Corollary 4.1 (Orbital stability). *Let $0 \leq \alpha < 4/\pi$, $\Lambda > 0$, $\nu \in \mathcal{C}$. Then*

1. *the set $\mathcal{M}$ of vacua minimizing $\mathcal{E}^\nu$ on the whole set $\mathcal{K}$ is orbitally stable in $\mathcal{K}$;*
2. *when $N \in \mathbb{Z}$ is such that (H_1) in Theorem 4.8 holds, the set $\mathcal{M}(N)$ of minimizers of $E^\nu(N)$ is orbitally stable in $\mathcal{K}(N)$.*

4.8 The positive temperature case

In a joint work [41] with Hainzl and Seiringer, we have partly extended all the above results to the non-zero temperature case, where minimizers behave in a rather different way. We briefly describe this now.

The starting point is as usual the Hartree-Fock free energy at temperature $T = \beta^{-1}$ which can be made rigorous in a box with periodic boundary conditions and a Fourier cut-off. It is expressed as

$$\mathcal{F}^\nu_{\mathrm{HF}}(\gamma) = \mathcal{E}^\nu_{\mathrm{HF}}(\gamma) - TS(\gamma) \tag{4.71}$$

where S is the entropy which reads for HF states

$$S(\gamma) = -\mathrm{Tr}\big((\gamma + 1/2)\log(\gamma + 1/2)\big) - \mathrm{Tr}\big((1/2 - \gamma)\log(1/2 - \gamma)\big).$$

We recall that $-1/2 \leq \gamma \leq 1/2$ is the renormalized density matrix of the HF state, i.e. $\gamma = P - 1/2$ where P is the usual one-body density matrix.

As before the first step is to construct the free vacuum state. It is defined as the thermodynamic limit of the free vacuum in a box of size L as $L \to \infty$. Similarly to Theorems 4.1, 4.2 and 4.3, it was proved in [41] that for $L \gg 1$, the free vacuum γ^0_L is unique and invariant by translations. It converges as $L \to \infty$ to a state γ^0 which satisfies the following self-consistent equation

$$\gamma^0 = \frac{1}{2}\left(\frac{1}{1 + e^{\beta D_{\gamma^0}}} - \frac{1}{1 + e^{-\beta D_{\gamma^0}}}\right)$$

where

$$D_\gamma = D^0 - \alpha\frac{(\gamma^0 - 1/2)(x, y)}{|x - y|}$$

is the usual mean-field operator. The limiting state γ^0 of the free vacuum is also the unique minimizer of the free energy per unit volume which reads

$$\mathcal{T}_T(\gamma) = \mathcal{T}(\gamma) + \frac{\beta^{-1}}{(2\pi)^3}\int_{B(0,\Lambda)} \mathrm{Tr}_{\mathbb{C}^4}\big[\left(\tfrac{1}{2} + \gamma(p)\right)\log\left(\tfrac{1}{2} + \gamma(p)\right) + \left(\tfrac{1}{2} - \gamma(p)\right)\log\left(\tfrac{1}{2} - \gamma(p)\right)\big]\,dp$$

---

[7] More precisely it is a consequence of the compactness of all minimizing sequences for $\mathcal{E}^\nu$ on $\mathcal{K}$ which is contained in the proof of Theorem 4.5.

where $\mathcal{T}$ is the energy per unit volume which was defined before in (4.15). Like for $T = 0$, it was shown in [41] that γ^0 takes the special form

$$\gamma^0(p) = f_1(|p|)\boldsymbol{\alpha} \cdot p + f_0(|p|)\beta \tag{4.72}$$

with $f_0, f_1 \leq 0$ a.e. on $B(0, \Lambda)$ and that D_{γ^0} satisfies

$$|D_{\gamma^0}| \geq |D^0|. \tag{4.73}$$

In particular we have that when $T > 0$

$$\sigma(\gamma^0) \subset \left[-\frac{1}{2} + \epsilon, -\epsilon\right] \cup \left[\epsilon, \frac{1}{2} - \epsilon\right] \tag{4.74}$$

for some $\epsilon > 0$. This can be seen from (4.73) and the fact that D_{γ^0} is a bounded operator on $\mathfrak{H}_\Lambda$ due to the presence of the ultraviolet cut-off. Notice also that we have formally $\rho_{\gamma^0} \equiv 0$ by (4.20), like for the $T = 0$ case. We assume that $T > 0$ henceforth.

We refer to [41] for a precise statement of the above claims. Again the uniqueness for L large is a non trivial result since the energy is not convex. The proof relies on a relative entropy estimate which we will explain with more details below.

As we did for the zero-temperature case, we now formally substract the free energy of the free vacuum to the formal free energy (4.71) and obtain the following expression:

$$\boxed{\mathcal{F}_T^\nu(Q) := TH(\gamma^0 + Q, \gamma^0) + \frac{\alpha}{2} D(\rho_Q - \nu, \rho_Q - \nu) - \frac{\alpha}{2} \iint_{\mathbb{R}^6} \frac{|Q(x,y)|^2}{|x-y|} dx\, dy} \tag{4.75}$$

where $Q = \gamma - \gamma^0$ and $H(\gamma, \gamma^0)$ is the *relative entropy* which is formally defined by the formula

$$TH(\gamma, \gamma^0) = \text{“}\mathrm{Tr}(D_{\gamma^0}(\gamma - \gamma^0)) - TS(\gamma) + TS(\gamma^0)\text{”}. \tag{4.76}$$

One can compute that

$$H(\gamma, \gamma^0) = \mathrm{Tr}\Big[\left(\tfrac{1}{2} + \gamma\right)\left(\ln\left(\tfrac{1}{2} + \gamma\right) - \ln\left(\tfrac{1}{2} + \gamma^0\right)\right) + \left(\tfrac{1}{2} - \gamma\right)\left(\ln\left(\tfrac{1}{2} - \gamma\right) - \ln\left(\tfrac{1}{2} - \gamma^0\right)\right)\Big]. \tag{4.77}$$

Note that when γ is a compact perturbation of γ^0 (as this will be the case in the following), we always have $\sigma_{\mathrm{ess}}(\gamma) = \sigma_{\mathrm{ess}}(\gamma^0)$. Hence $\sigma(\gamma)$ only contains eigenvalues of finite multiplicity in the neighborhood of ±1/2. Using an integral formula we easily see that Eq. (4.77) is well defined as soon as $\gamma - \gamma^0 \in \mathfrak{S}_1(\mathfrak{H}_\Lambda)$, since the spectrum of γ^0 does not contain $\pm 1/2$. When $Q = \gamma - \gamma^0$ is merely Hilbert-Schmidt, we may define the relative entropy by the integral formula

$$\boxed{H(\gamma, \gamma^0) = \mathrm{Tr}\left(\int_{-1}^{1} \frac{2}{1 + 2u\gamma^0}(\gamma - \gamma^0)\frac{1 - |u|}{1 + 2u\gamma}(\gamma - \gamma^0)\frac{1}{1 + 2u\gamma^0} du\right)} \tag{4.78}$$

It is clear that this provides a well defined object as one has

$$\forall \gamma \in \mathcal{K},\ \forall u \in [-1, 1], \quad 0 \leq \frac{1 - |u|}{1 + 2u\gamma} \leq 1 \quad \text{and} \quad 0 \leq \frac{1}{1 + 2u\gamma^0} \leq \frac{1}{\epsilon}$$

for some $\epsilon > 0$, by (4.74). It is not difficult to see that (4.78) and (4.77) coincide when $\gamma - \gamma^0 \in \mathfrak{S}_1(\mathfrak{H}_\Lambda)$. An important result proved in [41] is the

Theorem 4.14 (Properties of Relative Entropy [41])**.** *The functional* $Q \mapsto H(\gamma^0 + Q, \gamma^0)$ *defined in (4.78) is strongly continuous on*

$$\mathcal{K}' := \{Q \in \mathfrak{S}_2(\mathfrak{H}_\Lambda) \mid -1/2 \leq \gamma^0 + Q \leq 1/2\}$$

for the topology of $\mathfrak{S}_2(\mathfrak{H}_\Lambda)$*. It is convex, hence weakly lower semi-continuous (wlsc). Moreover, it is coercive on* $\mathcal{K}'$ *for the Hilbert-Schmidt norm:*

$$\forall \gamma \in \mathcal{K}, \qquad TH(\gamma^0 + Q, \gamma^0) \geq \operatorname{Tr}\left(|D^0|Q^2\right) \tag{4.79}$$

where we recall that $T = \beta^{-1}$ *is the temperature.*

We have already seen in the $T = 0$ case that the kinetic energy is controlling the norm of $\mathfrak{S}_1^{\mathcal{P}^0_-}(\mathfrak{H}_\Lambda)$. In the positive temperature case, the kinetic energy *only controls the Hilbert-Schmidt norm of* Q; it does not control the trace norm of Q^{++} and Q^{--} which could in principle only be in $\mathfrak{S}_2(\mathfrak{H}_\Lambda)$.

Equation (4.79) implies that the BDF free energy is bounded below:

$$\mathcal{F}_T^\nu(Q) \geq -\frac{\alpha}{2} D(\nu, \nu)$$

by the same proof as the one for $T = 0$. In particular, we obtain that $Q = 0$ is the unique minimizer when $\nu = 0$. This property can be transferred to a large box to prove the uniqueness when $L \gg 1$.

The free energy being bounded-below, it is natural to try to minimize it, like for $T = 0$. Unfortunately, we were not able to prove the existence of a minimizer when the exchange term is taken into account. However it was shown in [41] that any global minimizer (polarized vacuum at temperature $T > 0$) solves the following self-consistent equation, written in terms of $\gamma = \gamma^0 + Q$,

$$\gamma = \frac{1}{2}\left(\frac{1}{1 + e^{\beta D_\gamma}} - \frac{1}{1 + e^{-\beta D_\gamma}}\right)$$

where

$$D_\gamma = D^0 + \alpha(\rho_Q - \nu) * |\cdot|^{-1} - \alpha \frac{(\gamma - 1/2)(x, y)}{|x - y|}.$$

The reduced model and Debye screening

As for $T = 0$ it is possible to prove much more for the reduced model consisting in neglecting the exchange term. The free vacuum is of course given by the following formula

$$\tilde{\gamma}^0 = \frac{1}{2}\left(\frac{1}{1 + e^{\beta D^0}} - \frac{1}{1 + e^{-\beta D^0}}\right)$$

and the reduced-BDF free energy is defined similarly as before by

$$\boxed{\mathcal{F}_{T,\mathrm{r}}^\nu(Q) := TH(\tilde{\gamma}^0 + Q, \tilde{\gamma}^0) + \frac{\alpha}{2} D(\rho_Q - \nu, \rho_Q - \nu)} \tag{4.80}$$

with

$$Q \in \mathcal{K}'_\mathrm{r} = \{Q \in \mathfrak{S}_2(\mathfrak{H}_\Lambda) \mid -1/2 \leq \tilde{\gamma}^0 + Q \leq 1/2 \text{ and } \rho_Q \in \mathcal{C}\}.$$

Theorem 4.15 (Minimizer for reduced model at $T > 0$ and Debye screening [41])**.** *Assume* $T > 0$, $\alpha \geq 0$, $\Lambda > 0$ *and* $\nu \in \mathcal{C}$*. Then* $\mathcal{F}_{T,\mathrm{r}}^\nu$ *satisfies*

$$\forall Q \in \mathcal{K}'_\mathrm{r}, \qquad \mathcal{F}_{T,\mathrm{r}}^\nu(Q) \geq -\frac{\alpha}{2} D(\nu, \nu) \tag{4.81}$$

hence it is bounded below on $\mathcal{K}'_{\rm r}$. It has a unique minimizer *Q on $\mathcal{K}'_{\rm r}$. The operator $\gamma = \gamma^0 + Q$ satisfies the self-consistent equation*

$$\begin{cases} \gamma = \dfrac{1}{2}\left(\dfrac{1}{1+e^{\beta D_\gamma}} - \dfrac{1}{1+e^{-\beta D_\gamma}}\right), \\ D_\gamma := D^0 + \alpha(\rho_Q - \nu) * |\cdot|^{-1}. \end{cases} \tag{4.82}$$

If moreover $\nu \in L^1(\mathbb{R}^3)$, then $Q \in \mathfrak{S}_1(\mathfrak{H}_\Lambda)$ and the associated density ρ_Q is an $L^1(\mathbb{R}^3)$ function which satisfies

$$\int_{\mathbb{R}^3} \rho_Q = \int_{\mathbb{R}^3} \nu \qquad \textit{and} \qquad \big(\rho_Q - \nu\big) * \frac{1}{|x|} \in L^1(\mathbb{R}^3). \tag{4.83}$$

This result shows that the particles arrange themselves such that the total effective potential $\big(\rho_Q - \nu\big) * 1/|x|$ has a decay much faster than $1/|x|$. This implies that the nuclear charge of the external sources is completely screened. Within non-relativistic fermionic plasma this effect is known as *Debye-screening*. Let us emphasize that in order to recover such a screening, it is essential to calculate the Gibbs-state in a self-consistent way.

As a conclusion we have seen that the non-zero temperature case is quite different from the case $T = 0$. On the first hand the relative entropy only controls the Hilbert-Schmidt norm of Q, hence one *a priori* does not know wether Q^{++} and Q^{--} are trace-class. However (at least for the reduced model but a similar result is expected for the BDF model), the minimizers are indeed always trace-class. This is because the virtual particles of the vacuum have so much freedom due to the temperature that they are allowed to completely screen the nucleus, hence the total electrostatic potential decays much faster than expected. This is completely different from the $T = 0$ case for which Q is never a trace-class operator, a fact that is linked to renormalization as we have explained.

References

[1] D. Atkinson, P. W. Johnson, and P. Maris, *Dynamical mass generation in three-dimensional QED: Improved vertex function*, Phys. Rev. D, 42 (1990), pp. 602–609.

[2] V. Bach, J. M. Barbaroux, B. Helffer, and H. Siedentop, *On the stability of the relativistic electron-positron field*, Commun. Math. Phys., 201 (1999), pp. 445–460.

[3] V. Bach, E. H. Lieb, M. Loss, and J. P. Solovej, *There are no unfilled shells in unrestricted Hartree-Fock theory*, Phys. Rev. Lett., 72 (1994), pp. 2981–2983.

[4] V. Bach, E. H. Lieb, and J. P. Solovej, *Generalized Hartree-Fock theory and the Hubbard model*, J. Statist. Phys., 76 (1994), pp. 3–89.

[5] J. D. Bjorken and S. D. Drell, *Relativistic quantum fields*, McGraw-Hill Book Co., New York, 1965.

[6] E. Borie and G. Rinker, *The energy levels of muonic atoms*, Rev. Mod. Phys., 54 (1982), pp. 67–118.

[7] A. Bove, G. Da Prato, and G. Fano, *An existence proof for the Hartree-Fock time-dependent problem with bounded two-body interaction*, Commun. Math. Phys., 37 (1974), pp. 183–191.

[8] ———, *On the Hartree-Fock time-dependent problem*, Commun. Math. Phys., 49 (1976), pp. 25–33.

[9] É. Cancès, A. Deleurence, and M. Lewin, *A new approach to the modelling of local defects in crystals: the reduced Hartree-Fock case*, Commun. Math. Phys., 281 (2008), pp. 129–177.

[10] ———, *Non-perturbative embedding of local defects in crystalline materials*, J. Phys.: Condens. Matter, 20 (2008), p. 294213.

[11] É. Cancès and C. Le Bris, *On the time-dependent Hartree-Fock equations coupled with a classical nuclear dynamics*, Math. Models Methods Appl. Sci., 9 (1999), pp. 963–990.

[12] T. Cazenave and P.-L. Lions, *Orbital stability of standing waves for some nonlinear Schrödinger equations*, Commun. Math. Phys., 85 (1982), pp. 549–561.

[13] J. M. Chadam, *The time-dependent Hartree-Fock equations with Coulomb two-body interaction*, Commun. Math. Phys., 46 (1976), pp. 99–104.

[14] J. M. Chadam and R. T. Glassey, *Global existence of solutions to the Cauchy problem for time-dependent Hartree equations*, J. Math. Phys., 16 (1975), pp. 1122–1130.

[15] P. Chaix and D. Iracane, *From quantum electrodynamics to mean field theory: I. The Bogoliubov-Dirac-Fock formalism*, J. Phys. B, 22 (1989), pp. 3791–3814.

[16] P. Chaix, D. Iracane, and P.-L. Lions, *From quantum electrodynamics to mean field theory: II. Variational stability of the vacuum of quantum electrodynamics in the mean-field approximation*, J. Phys. B, 22 (1989), pp. 3815–3828.

[17] P. Dirac, *The quantum theory of the electron. II*, Proceedings Royal Soc. London (A), 118 (1928), pp. 351–361.

[18] ———, *A theory of electrons and protons*, Proceedings Royal Soc. London (A), 126 (1930), pp. 360–365.

[19] ———, *Discussion of the infinite distribution of electrons in the theory of the positron*, Proc. Camb. Philos. Soc., 30 (1934), pp. 150–163.

[20] ———, *Théorie du positron*, Solvay report, XXV (1934), pp. 203–212.

[21] F. J. Dyson, *The radiation theories of Tomonaga, Schwinger, and Feynman*, Phys. Rev. (2), 75 (1949), pp. 486–502.

[22] ———, *The S matrix in quantum electrodynamics*, Phys. Rev. (2), 75 (1949), pp. 1736–1755.

[23] ———, *Divergence of Perturbation Theory in Quantum Electrodynamics*, Phys. Rev., 85 (1952), pp. 631–632.

[24] E. Engel, *Relativistic Density Functional Theory: Foundations and Basic Formalism*, vol. 'Relativistic Electronic Structure Theory, Part 1.Fundamentals', Elsevier (Amsterdam), Schwerdtfeger, P. ed., 2002, ch. 10, pp. 524–624.

[25] E. Engel and R. M. Dreizler, *Field-theoretical approach to a relativistic Thomas-Fermi-Dirac-Weizsäcker model*, Phys. Rev. A, 35 (1987), pp. 3607–3618.

[26] M. Escobedo and L. Vega, *A semilinear Dirac equation in $H^s(\mathbf{R}^3)$ for $s > 1$*, SIAM J. Math. Anal., 28 (1997), pp. 338–362.

[27] M. J. Esteban, M. Lewin, and É. Séré, *Variational methods in relativistic quantum mechanics*, Bull. Amer. Math. Soc. (N.S.), 45 (2008), pp. 535–593.

[28] M. J. Esteban and É. Séré, *Solutions of the Dirac-Fock equations for atoms and molecules*, Commun. Math. Phys., 203 (1999), pp. 499–530.

[29] ———, *Nonrelativistic limit of the Dirac-Fock equations*, Ann. Henri Poincaré, 2 (2001), pp. 941–961.

[30] M. Flato, J. Simon, and E. Taflin, *On global solutions of the Maxwell-Dirac equations*, Commun. Math. Phys., 112 (1987), pp. 21–49.

[31] M. Flato, J. C. H. Simon, and E. Taflin, *Asymptotic completeness, global existence and the infrared problem for the Maxwell-Dirac equations*, Mem. Amer. Math. Soc., 127 (1997), pp. x+311.

[32] L. L. Foldy and E. Eriksen, *Some physical consequences of vacuum polarization*, Phys. Rev., II. Ser., 95 (1954), pp. 1048–1051.

[33] W. Furry, *A symmetry theorem in the positron theory*, Phys. Rev., 51 (1937), p. 125.

[34] W. Furry and J. Oppenheimer, *On the theory of the electron and positive*, Phys. Rev. (2), 45 (1934), pp. 245–262.

[35] V. Georgiev, *Small amplitude solutions of the Maxwell-Dirac equations*, Indiana Univ. Math. J., 40 (1991), pp. 845–883.

[36] R. Glauber, W. Rarita, and P. Schwed, *Vacuum polarization effects on energy levels in μ-mesonic atoms*, Phys. Rev., II. Ser., 120 (1960), pp. 609–613.

[37] P. Gravejat, M. Lewin, and É. Séré, *Ground state and charge renormalization in a nonlinear model of relativistic atoms*, Commun. Math. Phys., 286 (2009), pp. 179–215.

[38] W. Greiner and J. Reinhardt, *Quantum electrodynamics of Strong Fields*, Rep. Prog. Phys., 40 (1977), pp. 219–295.

[39] L. Gross, *The Cauchy problem for the coupled Maxwell and Dirac equations*, Comm. Pure Appl. Math., 19 (1966), pp. 1–15.

[40] R. Haag and T. A. Maris, *Dilatationally Invariant Quantum Electrodynamics of Electrons and Muons*, Physical Review, 132 (1963), pp. 2325–2330.

[41] C. Hainzl, M. Lewin, and R. Seiringer, *A nonlinear model for relativistic electrons at positive temperature*, Rev. Math. Phys., 20 (2008), pp. 1283 –1307.

[42] C. Hainzl, M. Lewin, and É. Séré, *Existence of a stable polarized vacuum in the Bogoliubov-Dirac-Fock approximation*, Commun. Math. Phys., 257 (2005), pp. 515–562.

[43] ———, *Self-consistent solution for the polarized vacuum in a no-photon QED model*, J. Phys. A, 38 (2005), pp. 4483–4499.

[44] ———, *Existence of atoms and molecules in the mean-field approximation of no-photon quantum electrodynamics*, Arch. Rational Mech. Anal., 192 (2009), pp. 453–499.

[45] C. Hainzl, M. Lewin, É. Séré, and J. P. Solovej, *A minimization method for relativistic electrons in a mean-field approximation of quantum electrodynamics*, Phys. Rev. A, 76 (2007), p. 052104.

[46] C. Hainzl, M. Lewin, and J. P. Solovej, *The mean-field approximation in quantum electrodynamics: the no-photon case*, Comm. Pure Appl. Math., 60 (2007), pp. 546–596.

[47] C. Hainzl, M. Lewin, and C. Sparber, *Existence of global-in-time solutions to a generalized Dirac-Fock type evolution equation*, Lett. Math. Phys., 72 (2005), pp. 99–113.

[48] W. Heisenberg, *Bemerkungen zur Diracschen Theorie des Positrons*, Z. Phys., 90 (1934), pp. 209–231.

[49] W. Heisenberg and H. Euler, *Folgerungen aus der Diracschen Theorie des Positrons*, Zeitschrift fur Physik, 98 (1936), pp. 714–732.

[50] W. Hunziker, *On the spectra of Schrödinger multiparticle Hamiltonians*, Helv. Phys. Acta, 39 (1966), pp. 451–462.

[51] C. Itzykson and J. B. Zuber, *Quantum field theory*, McGraw-Hill International Book Co., New York, 1980. International Series in Pure and Applied Physics.

[52] K. Johnson, M. Baker, and R. Willey, *Self-Energy of the Electron*, Physical Review, 136 (1964), pp. 1111–1119.

[53] G. Källén, *On the definition of the renormalization constants in quantum electrodynamics*, Helvetica Phys. Acta, 25 (1952), pp. 417–434.

[54] T. Kato, *Perturbation theory for linear operators*, Springer, second ed., 1995.

[55] L. Landau and I. Pomerančuk, *On point interaction in quantum electrodynamics*, Dokl. Akad. Nauk SSSR (N.S.), 102 (1955), pp. 489–492.

[56] L. D. Landau, *On the quantum theory of fields*, in Niels Bohr and the development of physics, McGraw-Hill Book Co., New York, N. Y., 1955, pp. 52–69.

[57] L. D. Landau, A. A. Abrikosov, and I. M. Halatnikov, *The mass of the electron in quantum electrodynamics*, Doklady Akad. Nauk SSSR (N.S.), 96 (1954), pp. 261–264.

[58] L. D. Landau, A. A. Abrikosov, and L. Halatnikov, *On the quantum theory of fields*, Nuovo Cimento, 3 (1956), pp. 80–104.

[59] H. Lehmann, *Über Eigenschaften von Ausbreitungsfunktionen und Renormierungskonstanten quantisierter Felder*, Nuovo Cimento, 11 (1954), pp. 342–357.

[60] E. H. Lieb, *Bound on the maximum negative ionization of atoms and molecules*, Phys. Rev. A, 29 (1984), pp. 3018–3028.

[61] E. H. Lieb and M. Loss, *Existence of atoms and molecules in non-relativistic quantum electrodynamics.*, Adv. Theor. Math. Phys., 7 (2003), pp. 667–710.

[62] E. H. Lieb and H. Siedentop, *Renormalization of the regularized relativistic electron-positron field*, Commun. Math. Phys., 213 (2000), pp. 673–683.

[63] E. H. Lieb and B. Simon, *The Hartree-Fock theory for Coulomb systems*, Commun. Math. Phys., 53 (1977), pp. 185–194.

[64] P.-L. Lions, *The concentration-compactness principle in the calculus of variations. The locally compact case, Part I*, Ann. Inst. Henri Poincaré, 1 (1984), pp. 109–149.

[65] ———, *Solutions of Hartree-Fock equations for Coulomb systems*, Commun. Math. Phys., 109 (1987), pp. 33–97.

[66] S. Machihara, K. Nakanishi, and T. Ozawa, *Small global solutions and the nonrelativistic limit for the nonlinear Dirac equation*, Rev. Mat. Iberoamericana, 19 (2003), pp. 179–194.

[67] M. Mittleman, *Theory of relativistic effects on atoms: Configuration-space hamiltonian*, Phys. Rev. A, 24 (1981), pp. 1167–1175.

[68] P. J. Mohr, G. Plunien, and G. Soff, *QED corrections in heavy atoms*, Phys. Rep., 293 (1998), pp. 227–369.

[69] G. Nenciu, *Existence of the spontaneous pair creation in the external field approximation of Q.E.D*, Commun. Math. Phys., 109 (1987), pp. 303–312.

[70] W. Pauli, *Relativistic field theories of elementary particles*, Rev. Modern Physics, 13 (1941), pp. 203–232.

[71] W. Pauli and M. Rose, *Remarks on the polarization effects in the positron theory*, Phys. Rev II, 49 (1936), pp. 462–465.

[72] P. Pickl and D. Durr, *On adiabatic pair creation*, Commun. Math. Phys., 282 (2008), pp. 161–198.

[73] I. Y. Pomeranchuk, V. V. Sudakov, and K. A. Ter-Martirosyan, *Vanishing of Renormalized Charges in Field Theories with Point Interaction*, Physical Review, 103 (1956), pp. 784–802.

[74] M. Reed and B. Simon, *Methods of Modern Mathematical Physics. I. Functionnal Analysis*, Academic Press, 1972.

[75] P.-G. Reinhard, W. Greiner, and H. Arenhövel, *Electrons in strong external fields*, Nuclear Physics A, 166 (1971), pp. 173–197.

[76] J. Reinhardt, B. Müller, and W. Greiner, *Theory of positron production in heavy-ion collisions*, Phys. Rev. A, 24 (1981), pp. 103–128.

[77] G. Scharf and H. P. Seipp, *Charged vacuum, spontaneous positron production and all that*, Physics Letters B, 108 (1982), pp. 196–198.

[78] J. Schwinger, *Quantum electrodynamics. I. A covariant formulation*, Phys. Rev. (2), 74 (1948), pp. 1439–1461.

[79] ———, *On the Green's functions of quantized fields. I, II*, Proc. Nat. Acad. Sci. U. S. A., 37 (1951), pp. 452–455, 455–459.

[80] R. Serber, *A Note on Positron Theory and Proper Energies*, Physical Review, 49 (1936), pp. 545–550.

[81] B. Simon, *Trace ideals and their applications*, vol. 35 of London Mathematical Society Lecture Note Series, Cambridge University Press, Cambridge, 1979.

[82] J. P. Solovej, *Proof of the ionization conjecture in a reduced Hartree-Fock model.*, Invent. Math., 104 (1991), pp. 291–311.

[83] ———, *The ionization conjecture in Hartree-Fock theory*, Ann. of Math. (2), 158 (2003), pp. 509–576.

[84] B. Swirles, *The relativistic self-consistent field.*, Proc. R. Soc. Lond., Ser. A, 152 (1935), pp. 625–649.

[85] N. Szpak, *On the problem of "spontaneous pair creation" in strong electric fields*, Nuclear Instruments and Methods in Physics Research B, 205 (2003), pp. 30–35.

[86] ———, *Spontaneous particle creation in time-dependent overcritical fields*, Journal of Physics A Mathematical General, 41 (2008), p. 164059.

[87] B. Thaller, *The Dirac equation*, Texts and Monographs in Physics, Springer-Verlag, Berlin, 1992.

[88] E. Uehling, *Polarization effects in the positron theory*, Phys. Rev. (2), 48 (1935), pp. 55–63.

[89] C. Van Winter, *Theory of finite systems of particles. I. The Green function*, Mat.-Fys. Skr. Danske Vid. Selsk., 2 (1964).

[90] G. M. Zhislin, *A study of the spectrum of the Schrödinger operator for a system of several particles*, Trudy Moskov. Mat. Obšč., 9 (1960), pp. 81–120.

5

Nonrelativistic crystal in the Hartree-Fock approximation

Describing the electronic state of crystals with local defects is a major issue in solid-state physics, materials science and nano-electronics [30, 21, 34]. In this chapter, we develop a theory based on formal analogies between the Fermi sea of a perturbed crystal and the polarized Dirac sea in Quantum Electrodynamics in the presence of an external electrostatic field, as described in Chapter 4. Using and adapting the methods presented before, we are able to propose a new mathematical approach for the self-consistent description of a crystal in the presence of local defects. We focus on the *Hartree-Fock* (HF) and on the *reduced Hartree-Fock* (rHF) models.

The chapter is organized as follows. In Section 5.1 we define the HF and rHF ground states of the *perfect* crystal and study its thermodynamic limit; we essentially complement results of Catto, Le Bris and Lions [13]. Section 5.2 is devoted to the definition and the study of the rHF model for the crystal in the presence of a defect. In some sense this is similar to the BDF model studied in the previous chapter, but with many technical and conceptual differences that will be emphasized. Eventually, we describe in Section 5.3 a numerical approach associated with our theoretical study and test it on a simple 1D system.

5.1 Perfect crystal

We start by studying the perfect crystal. We define two models, the Hartree-Fock and the reduced Hartree-Fock, following [13]. We prove the existence of minimizers and give some of their properties. We also show in Section 5.1.3 that the reduced Hartree-Fock Fermi sea is actually the correct state as it can be obtained by a thermodynamic limit (in the Hartree-Fock case this is not known).

To simplify the mathematical formulas, we will not explicitly take the spin variable into account and we will assume (except in Section 5.2.4) that the host crystal is cubic with a single atom of charge Z per unit cell. The arguments below can be easily extended to the general case. In the whole chapter, we denote by $\mathcal{R} := \mathbb{Z}^3$ the lattice on which are placed the nuclei, and by $\Gamma := [-1/2, 1/2)^3$ the corresponding unit cell. We denote by $\mathcal{R}^* := 2\pi\mathbb{Z}^3$ the associated dual lattice and by $\Gamma^* := [-\pi/2, \pi/2)^3$ the so-called *Brillouin zone.*

5.1.1 The periodic Coulomb interaction

In this paragraph, we introduce two functions G and W which we shall need throughtout the chapter. They will respectively yield the so-called direct and exchange terms of periodic Hartree-

Fock theory. We start by introducing the $\mathbb{Z}^3$-periodic Green kernel of the Poisson interaction, denoted by G and uniquely defined by

$$\begin{cases} -\Delta G = 4\pi \left(\sum_{k\in\mathbb{Z}^3} \delta_k - 1 \right) \\ G\ \mathbb{Z}^3\text{-periodic},\ \min_{\mathbb{R}^3} G = 0 \end{cases}$$

and where the first equation holds in the distributional sense. The Fourier expansion of G is

$$G(x) = h + \sum_{k\in 2\pi\mathbb{Z}^3\setminus\{0\}} \frac{4\pi}{|k|^2} e^{ik\cdot x} \tag{5.1}$$

with $h = \int_\Gamma G > 0$. The electrostatic potential associated with a $\mathbb{Z}^3$-periodic density $\rho \in L^1_{\rm loc}(\mathbb{R}^3) \cap L^3_{\rm loc}(\mathbb{R}^3)$ is the $\mathbb{Z}^3$-periodic function defined as

$$(\rho *_{\mathcal{R}} G)(x) := \int_\Gamma G(x-y)\,\rho(y)\,dy.$$

We also set for any $\mathbb{Z}^3$-periodic functions f and g

$$D_G(f,g) := \int_\Gamma \int_\Gamma G(x-y)\,f(x)\,g(y) dx\,dy.$$

Next we introduce the following function [13]

$$W(\eta, z) = \sum_{k\in\mathbb{Z}^3} \frac{e^{ik\cdot\eta}}{|z+k|}, \quad \eta, z \in \mathbb{R}^3. \tag{5.2}$$

The function $e^{i\eta\cdot x} W(\eta, x)$ is Γ-periodic with respect to x, when η is fixed. So we can write W as a Fourier series and obtain

$$W(\eta, x) = 4\pi e^{-i\eta\cdot x} \sum_{k\in 2\pi\mathbb{Z}^3} \frac{e^{ik\cdot x}}{|\eta - k|^2}. \tag{5.3}$$

5.1.2 The (reduced) periodic Hartree-Fock functional

We now define the *periodic Hartree-Fock* functional (and its corresponding *reduced* version) which were studied before in [13]. The main object of interest will, as usual, be the so-called (periodic) density matrix of the electrons. We define the translation operator τ_k acting on $L^2_{\rm loc}(\mathbb{R}^3)$ as follows: $\tau_k u(x) = u(x-k)$ and introduce the following variational set of density matrices:

$$\mathcal{P}_{\rm per} = \Big\{ \gamma \in S\big(L^2(\mathbb{R}^3)\big) \mid 0 \leq \gamma \leq 1,\ \forall k \in \mathbb{Z}^3,\ \tau_k\gamma = \gamma\tau_k, \\ \int_{\Gamma^*} {\rm Tr}_{L^2_\xi(\Gamma)}((1-\Delta_\xi)^{1/2}\gamma_\xi(1-\Delta_\xi)^{1/2})\,d\xi < \infty \Big\}.$$

In the whole paper, we use the notation $(A_\xi)_{\xi\in\Gamma^*}$ for the Bloch waves decomposition of a periodic operator A, see [31, 13]:

$$A = \frac{1}{(2\pi)^3} \int_{\Gamma^*} A_\xi\,d\xi, \qquad A_\xi \in \mathcal{S}(L^2_\xi(\Gamma)),$$

$$L^2_\xi(\Gamma) = \big\{ u \in L^2_{\rm loc}(\mathbb{R}^3) \mid \tau_k u = e^{-ik\cdot\xi} u,\ \forall k \in \mathbb{Z}^3 \big\}$$

which corresponds to the decomposition in fibers $L^2(\mathbb{R}^3)=\int_{\Gamma^*}^{\oplus} d\xi L^2_\xi(\Gamma)$.

For any $\gamma\in\mathcal{P}_{\rm per}$ (and almost every $\xi\in\Gamma^*$), we denote by $\gamma_\xi(x,y)$ the integral kernel of $\gamma_\xi\in\mathfrak{S}_1(L^2_\xi(\Gamma))$. The density of γ is then the non-negative $\mathbb{Z}^3$-periodic function of $L^1_{\rm loc}(\mathbb{R}^3)\cap L^3_{\rm loc}(\mathbb{R}^3)$ defined as

$$\rho_\gamma(x):=\frac{1}{(2\pi)^3}\int_{\Gamma^*}\gamma_\xi(x,x)\,d\xi.$$

Later we shall add the constraint that the system is neutral and restrict to states $\gamma\in\mathcal{P}_{\rm per}$ satisfying $\int_\Gamma\rho_\gamma(x)dx=Z$ where Z is the charge of the only nucleus in each unit cell.

The *periodic Hartree-Fock* and *reduced Hartree-Fock functional* give, for a system in a state $\gamma\in\mathcal{P}_{\rm per}$ the energy per unit volume in the corresponding model. They are respectively defined by

$$\boxed{\mathcal{E}^{\rm per}_{\rm HF}(\gamma):=\int\limits_{\Gamma^*}{\rm Tr}_{L^2_\xi(\Gamma)}\left(-\frac12\Delta_\xi\gamma_\xi\right)\frac{d\xi}{(2\pi)^3}+\frac12 D_G(\rho_\gamma-\mu_{\rm per},\rho_\gamma-\mu_{\rm per})-\frac12X(\gamma,\gamma)}\tag{5.4}$$

and

$$\boxed{\mathcal{E}^{\rm per}_{\rm rHF}(\gamma):=\int\limits_{\Gamma^*}{\rm Tr}_{L^2_\xi(\Gamma)}\left(-\frac12\Delta_\xi\gamma_\xi\right)\frac{d\xi}{(2\pi)^3}+\frac12 D_G(\rho_\gamma-\mu_{\rm per},\rho_\gamma-\mu_{\rm per}),}\tag{5.5}$$

for any $\gamma\in\mathcal{P}_{\rm per}$. In the above formula, the exchange term $X(\gamma,\gamma)/2$ is defined for any $\beta,\gamma\in\mathcal{P}_{\rm per}$ as

$$X(\beta,\gamma)=\frac{1}{(2\pi)^6}\iint\limits_{\Gamma^*\times\Gamma^*}d\xi\,d\xi'\iint\limits_{\Gamma\times\Gamma}dx\,dy\;\beta(\xi,x,y)W(\xi-\xi',x-y)\overline{\gamma(\xi',x,y)}\tag{5.6}$$

where W is the function defined in (5.2). We remark that $\overline{W(-\eta,z)}=W(\eta,z)=\overline{W(\eta,-z)}$, so $X(\beta,\gamma)=\overline{X(\gamma,\beta)}=X(\gamma,\beta)$. Lastly, the periodic background density of the nuclei is given by

$$\mu_{\rm per}:=Z\sum_{k\in\mathbb{Z}^3}\tau_k m.\tag{5.7}$$

We assume that either $m=\delta_0$ (pointwise nuclei), or that m is a $C_0^\infty(\mathbb{R}^3)$ nonnegative function such that $\int_{\mathbb{R}^3}m(x)=1$ and with a support small enough such that all the $\tau_k m$ have disjoint supports (extended nuclei).

The existence of a minimizer for $\mathcal{E}^{\rm per}_{\rm HF}$ and $\mathcal{E}^{\rm per}_{\rm rHF}$ on the set of density matrices $\gamma\in\mathcal{P}_{\rm per}$ such that $\int_\gamma\rho_\gamma=Z$ was proved by Catto, Le Bris and Lions in [13], see Theorem 2.1 p. 696 and Theorem 2.3 p. 698. However, the form of the self-consistent equation satisfied by a minimizer as well as the uniqueness in the reduced case were missing. These open problems were solved with A. Deleurence and É. Cancès in [9] for the rHF case and with M. Ghimenti in [17] for the HF case:

Theorem 5.1 (Minimizers in the periodic reduced Hartree-Fock case [9])**.** *Let $Z\in\mathbb{N}\setminus\{0\}$. The minimization problem*

$$\boxed{I^{\rm per}_{\rm rHF}=\inf_{\substack{\gamma\in\mathcal{P}_{\rm per},\\ \int_\Gamma\rho_\gamma=Z}}\mathcal{E}^{\rm per}_{\rm rHF}(\gamma)}\tag{5.8}$$

admits a unique minimizer $\gamma^0_{\rm per}$. *Denoting by*

$$H^0_{\rm per}:=-\frac{\Delta}{2}+(\rho_{\gamma^0_{\rm per}}-\mu_{\rm per})*_{\mathcal{R}}G,\tag{5.9}$$

the corresponding periodic mean-field Hamiltonian, $\gamma^0_{\rm per}$ is solution to the following nonlinear equation

$$\gamma^0_{\rm per}=\chi_{(-\infty,\epsilon_F]}(H^0_{\rm per}),\tag{5.10}$$

where ϵ_F *is a Lagrange multiplier called* Fermi level*, which can be interpreted as a chemical potential. Additionally, for any* $\epsilon_F \in \mathbb{R}$ *such that* (5.10) *holds,* $\gamma^0_{\rm per}$ *is the unique minimizer on* $\mathcal{P}_{\rm per}$ *of the energy functional*

$$\gamma \mapsto \mathcal{E}^{\rm per}_{\rm rHF}(\gamma) - \epsilon_F \int_\Gamma \rho_\gamma.$$

Theorem 5.1 contains three main results that were not present in [13]: first $\gamma^0_{\rm per}$ is unique, second it is a projector, and third it satisfies Equation (5.10). These three properties are crucial for a proper construction of the model for the crystal with a defect, as will be done below in Section 5.2. It can easily be seen that $(\rho_{\gamma^0_{\rm per}} - \mu_{\rm per}) *_{\mathcal{R}} G$ belongs to $L^2_{\rm loc}(\mathbb{R}^3)$. By a result of Thomas [35] this implies that the spectrum of $H^0_{\rm per}$ is purely absolutely continuous. Hence, we must have $\ker(H^0_{\rm per} - \epsilon_{\rm F}) = \{0\}$ which easily implies the uniqueness of $\gamma^0_{\rm per}$, see the Appendix of [9].

In the Hartree-Fock case, the following was proved in [17]:

Theorem 5.2 (Minimizers in the periodic Hartree-Fock case [17])**.** *Let* $Z \in \mathbb{N}\setminus\{0\}$*. Assume that* γ *is a minimizer of*

$$\boxed{I^{\rm per}_{\rm HF} = \inf_{\substack{\gamma\in\mathcal{P}_{\rm per},\\ \int_\Gamma \rho_\gamma = Z}} \mathcal{E}^{\rm per}_{\rm HF}(\gamma)} \tag{5.11}$$

and denote by

$$(H_\gamma)_\xi = -\frac{\Delta_\xi}{2} + (\rho_\gamma - \mu_{\rm per}) *_{\mathcal{R}} G - (2\pi)^{-3}\int_{\Gamma^*} W(\xi'-\xi, x-y)\gamma_{\xi'}(x,y)\,d\xi' \tag{5.12}$$

the Bloch transform of the associated periodic Hartree-Fock mean-field Hamiltonian. Then γ *solves the following nonlinear equation:*

$$\gamma = \chi_{(-\infty,\epsilon_{\rm F})}(H_\gamma) + s\chi_{\{\epsilon_{\rm F}\}}(H_\gamma), , \tag{5.13}$$

where ϵ_F *is a Lagrange multiplier called* Fermi level*, which can be interpreted as a chemical potential, and* $s \in \{0,1\}$*.*

As usual it is not expected that minimizers will always be unique in the Hartree-Fock case. When the exchange term is kept in the model as in Theorem 5.2, the situation is much more complicated than in the reduced case. The main difficulty is that the Bloch decomposition of the last term of H_γ depends in a non trivial way of ξ. In particular, we do not know if the spectrum of H_γ is purely absolutely continuous. It is in principle possible that $\epsilon_{\rm F}$ is an eigenvalue (of infinite multiplicity) of H_γ in (5.13). However, we are able to prove that any minimizer is automatically a projector. Theorem 5.2 even states that either the minimizer γ does not contain the eigenspace corresponding to the eigenvalue $\epsilon_{\rm F}$ ($s = 0$) or it fills it completely ($s = 1$). Equation (5.13) as well as the fact that γ is a projector were absent in [13].

The idea of the proof is somewhat similar to that of the atomic case [24, 3, 4, 5, 25]. We transfer within the last level $\epsilon_{\rm F}$ some mass from an eigenvector to another one and show that the energy must decrease. The originality of the periodic case studied here is that the transfer needs to be done between two different Bloch sectors $L^2_{\xi_1}(\Gamma)$ and $L^2_{\xi_2}(\Gamma)$.

In the rest of the chapter, we will focus on the simpler *reduced* case. We will often need to know that the perfect crystal is an *insulator* or a *semi-conductor.* This means that there is a gap between the last filled band and the first unfilled band. When necessary, we shall hence make the following assumption (recall that Z is the total charge of the nuclei in each cell):

(A1) *There is a gap between the* Z*-th and the* $(Z+1)$*-st bands, i.e.* $\Sigma^+_Z < \Sigma^-_{Z+1}$*, where* Σ^+_Z *and* Σ^-_{Z+1} *are respectively the maximum and the minimum of the* Z*-th and the* $(Z+1)$*-st bands of* $H^0_{\rm per}$*.*

We emphasize that Assumption **(A1)** is a condition on the solution $\gamma^0_{\rm per}$ of the *nonlinear* problem (5.10). Note that under **(A1)**, one has $\gamma^0_{\rm per} = \chi_{(-\infty,\epsilon_F]}(H^0_{\rm per})$ for any $\epsilon_F \in (\Sigma^+_Z, \Sigma^-_{Z+1})$.

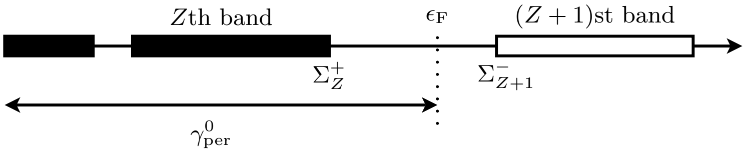


Figure 5.1: Spectrum of $H^0_{\rm per}$.

5.1.3 Thermodynamic limit of the supercell model in the reduced case

In the previous section, we have introduced two Hartree-Fock type models for a periodic crystal but we have not justified these definitions. Indeed, it was proved in [13] that the periodic reduced HF model is the thermodynamic limit of the corresponding theory for finite systems, as studied in Chapter 3. This result was itself complemented in [9], by (vaguely speaking) adding periodic boundary conditions on the sequence of domains considered in the limit, as will be explained in details in this section.

For the full Hartree-Fock theory with exchange term, it is known from Chapter 3 that the thermodynamic limit exists, but the limit could be different from the model defined in the previous section, based on the energy $\mathcal{E}^{\rm per}_{\rm HF}$. In principle symmetry breaking could occur in such a way that a minimizer would have another periodicity than the one given from $\mathbb{Z}^3$. In particular it could be that the HF ground state is actually $k\mathbb{Z}^3$-periodic, for some integer k. It would then be described by the same model as in the previous section, with $\mathbb{Z}^3$ replaced by $k\mathbb{Z}^3$. In practice it is however almost always assumed in quantum chemistry calculations that the Hartree-Fock state is $\mathbb{Z}^3$ periodic, leading to the energy $\mathcal{E}^{\rm per}_{\rm HF}$ that we have introduced.

In [13], the authors consider a domain Ω_L, and assume that the nuclei are located on $\mathbb{Z}^3\cap\Omega_L$. Then they consider the rHF model for N electrons living *in the whole space*, with $N = Z|\Omega_L|$ chosen to impose neutrality. Denoting by ρ_L the ground state electronic density of the latter problem, it is proved in [13, Thm 2.2] that the energy per unit volume converges to $I^{\rm per}_{\rm rHF}$ when the sequence Ω_L grows (in some appropriate sense similar to what we have done in Chapter 3), and that the following holds:

$$\sqrt{L^{-3}\sum_{k\in\mathbb{Z}^3\cap\Omega_L}\rho_L(x-k)}\to\sqrt{\rho_{\gamma^0_{\rm per}}} \tag{5.14}$$

weakly in $H^1_{\rm loc}(\mathbb{R}^3)$, strongly in $L^p_{\rm loc}(\mathbb{R}^3)$ for all $2\le p<6$ and almost everywhere on $\mathbb{R}^3$ when $L\to\infty$.

Another way for performing thermodynamic limits is to confine the nuclei *and* the electrons in the domain Ω_L by means of Dirichlet boundary conditions for the electrons, as explained in Chapter 3. Note that confining the electrons in Ω_L is mandatory in the positive temperature case.

Another possibility, perhaps less satisfactory from a physical viewpoint but more directly related to practical calculations (see e.g. [14]), is to take $\Omega_L = \Lambda_L := [-L/2, L/2)^3$ (with $L\in\mathbb{N}$) and to impose periodic boundary conditions on the box Λ_L. Usually the Coulomb interaction is also replaced by a $(L\mathbb{Z}^3)$-periodic Coulomb potential, leading to the so-called *supercell model* which will be described in detail below. This approach has the advantage of respecting the symmetry of the system in the crystal case. It was used by Hainzl, Lewin and Solovej in [18] to justify the Hartree-Fock approximation of no-photon Quantum Electrodynamics, as explained in Chapter 4, Theorem 4.3.

Of course the conjecture is that the final results (the energy per unit cell and the ground state density of the crystal) should not depend on the chosen thermodynamic limit procedure. This is actually the case for the reduced Hartree-Fock model of the crystal.

Let us now describe the supercell model. For $L\in\mathbb{N}\setminus\{0\}$, we introduce the supercell $\Lambda_L = [-L/2, L/2)^3$ and the Hilbert space

$$L^2_{\rm per}(\Lambda_L) = \left\{\varphi\in L^2_{\rm loc}(\mathbb{R}^3)\ |\ \varphi\ (L\mathbb{Z}^3)\text{-periodic}\right\}.$$

We also introduce the $L\mathbb{Z}^3$-periodic Coulomb potential G_L defined as the unique solution to

$$\begin{cases} -\Delta G_L = 4\pi \left(\displaystyle\sum_{k \in L\mathbb{Z}^3} \delta_k - \frac{1}{L^3} \right) \\ G_L \ L\mathbb{Z}^3\text{-periodic}, \quad \min_{\mathbb{R}^3} G_L = 0. \end{cases}$$

An admissible electronic state is then described by a one-body density matrix γ in

$$\mathcal{P}_{\mathrm{sc},L} = \left\{ \gamma \in \mathfrak{S}_1(L^2_{\mathrm{per}}(\Lambda_L)) \mid \gamma^* = \gamma, \ 0 \le \gamma \le 1, \ \mathrm{Tr}_{L^2_{\mathrm{per}}(\Lambda_L)}(-\Delta\gamma) < +\infty \right\}.$$

Throughout this section, we use the subscript 'sc' to indicate that we consider the thermodynamic limit of the supercell model. The reduced Hartree-Fock energy functional of the supercell model is defined for $\gamma \in \mathcal{P}_{\mathrm{sc},L}$ as

$$\boxed{\mathcal{E}^0_{\mathrm{sc},L}(\gamma) = \mathrm{Tr}_{L^2_{\mathrm{per}}(\Lambda_L)} \left(-\frac{1}{2}\Delta\gamma \right) + \frac{1}{2} \int_{\Lambda_L} \int_{\Lambda_L} G_L(x-y) \, (\rho_\gamma - \mu_{\mathrm{per}})(x) \, (\rho_\gamma - \mu_{\mathrm{per}})(y) \, dx \, dy}$$

where we recall that $\mu_{\mathrm{per}}(x) = \sum_{R \in \mathbb{Z}^3} Zm(x-R)$ is a $\mathbb{Z}^3$- (thus $L\mathbb{Z}^3$-) periodic function. The reduced Hartree-Fock ground state energy for a neutral system in the box of size L is then given by

$$\boxed{I^0_{\mathrm{sc},L} = \inf_{\substack{\gamma \in \mathcal{P}_{\mathrm{sc},L}, \\ \int_{\Lambda_L} \rho_\gamma = ZL^3}} \mathcal{E}^0_{\mathrm{sc},L}(\gamma)} \tag{5.15}$$

Let us recall that $I^{\mathrm{per}}_{\mathrm{rHF}}$, γ^0_{per} and H^0_{per} are defined in Section 5.1.2. The following was proved in [9]:

Theorem 5.3 (Thermodynamic limit of the defect-free supercell model [9])**.** *Let $Z > 0$.*

i) For all $L \in \mathbb{N} \setminus \{0\}$, the minimizing problem $I^0_{\mathrm{sc},L}$ has at least one minimizer, and all the minimizers share the same density. This density is $\mathbb{Z}^3$-periodic. Besides, there is one minimizer $\gamma^0_{\mathrm{sc},L}$ of (5.15) *which commutes with the translations τ_k, $k \in \mathbb{Z}^3$.*

ii) The following thermodynamic limit properties hold true:

- (Convergence of the energy per unit cell).

$$\lim_{L \to \infty} \frac{I^0_{\mathrm{sc},L}}{L^3} = I^{\mathrm{per}}_{\mathrm{rHF}};$$

- (Convergence of the density).

$$\sqrt{\rho_{\gamma^0_{\mathrm{sc},L}}} \rightharpoonup \sqrt{\rho_{\gamma^0_{\mathrm{per}}}} \quad \textit{weakly in } H^1_{\mathrm{loc}}(\mathbb{R}^3), \tag{5.16}$$

$$\rho_{\gamma^0_{\mathrm{sc},L}} \to \rho_{\gamma^0_{\mathrm{per}}} \quad \textit{strongly in } L^p_{\mathrm{loc}}(\mathbb{R}^3) \ \textit{ for } 1 \le p < 3 \textit{ and a.e.};$$

- (Convergence of the mean-field Hamiltonian and its spectrum). *Let*

$$H^0_{\mathrm{sc},L} = -\frac{\Delta}{2} + (\rho_{\gamma^0_{\mathrm{sc},L}} - \mu_{\mathrm{per}}) *_{\mathcal{R}} G$$

seen as an operator acting on $L^2(\mathbb{R}^3)$. Then, for all $L \in \mathbb{N} \setminus \{0\}$, $H^0_{\mathrm{sc},L} - H^0_{\mathrm{per}}$ is a bounded operator and

$$\lim_{L \to \infty} \left\| H^0_{\mathrm{sc},L} - H^0_{\mathrm{per}} \right\| = 0.$$

Denoting by $(\lambda^L_n(\xi))_{n \in \mathbb{N} \setminus \{0\}}$ the nondecreasing sequence of eigenvalues of $(H^0_{\mathrm{sc},L})_\xi$ for $\xi \in \Gamma^*$, one has

$$\lim_{L \to \infty} \sup_{n \in \mathbb{N} \setminus \{0\}} \sup_{\xi \in \Gamma^*} \left| \lambda^L_n(\xi) - \lambda_n(\xi) \right| = 0 \tag{5.17}$$

where $(\lambda_n(\xi))_{n\geq 1}$ *are the eigenvalues of* $(H^0_{\rm per})_\xi$ *introduced in Theorem 5.1.*

iii) Assume in addition that **(A1)** *holds. Fix some* $\epsilon_F\in(\Sigma_Z^+,\Sigma_{Z+1}^-)$*. Then for* L *large enough, the minimizer* $\gamma^0_{{\rm sc},L}$ *of* $I^0_{{\rm sc},L}$ *is unique. It is also the unique minimizer of the following problem*

$$I^0_{{\rm sc},L,\epsilon_F}:=\inf\left\{\mathcal{E}^0_{{\rm sc},L}(\gamma)-\epsilon_F{\rm Tr}_{L^2_{\rm per}(\Lambda_L)}(\gamma),\quad \gamma\in\mathcal{P}_{{\rm sc},L}\right\}. \qquad (5.18)$$

Notice that some of the above assertions are more precise for the supercell model than for the thermodynamic limit procedure considered in [13, Thm 2.2] (compare for instance (5.16) with (5.14)). This is because the supercell model respects the symmetry of the system, allowing in particular to have a minimizer $\gamma^0_{{\rm sc},L}$ in the box of size L^3 which is periodic for the lattice $\mathbb{Z}^3$. For an insulator, the uniqueness of $\gamma^0_{{\rm sc},L}$ for large L and the convergence properties of *ii)* are also very interesting for computational purposes.

5.2 Crystal with a defect

5.2.1 The reduced-Hartree-Fock energy with a defect

We now define the reduced Hartree-Fock model describing the behavior of the Fermi sea and possibly of a finite number of bound electrons (or holes) close to a local defect. Our model is an obvious transposition of the Bogoliubov-Dirac-Fock model which was presented before in Chapter 4.

Assume that the periodic nuclear density $\mu_{\rm per}$ defined in (5.7) is replaced by a locally perturbed nuclear density $\mu_{\rm per}+\nu$. The defect ν can model a vacancy, an interstitial atom, or an impurity, with possible local rearrangement of the neighboring atoms. The main idea underlying the model is to define a *finite* energy by subtracting the *infinite* energy of the periodic Fermi sea $\gamma^0_{\rm per}$, from the *infinite* energy of the perturbed system under consideration. This is of course exactly the same method as the one of Chapter 4 for relativistic systems. Formally, one obtains for a test state γ

$$\begin{aligned}\mathcal{E}^{\rm rHF}_{\mu_{\rm per}+\nu}(\gamma)-\mathcal{E}^{\rm rHF}_{\mu_{\rm per}+\nu}(\gamma^0_{\rm per})\text{ `` = '' }{\rm Tr}\left(H^0_{\rm per}(\gamma-\gamma^0_{\rm per})\right)&\\ -\int_{\mathbb{R}^3}\int_{\mathbb{R}^3}\frac{\nu(x)\rho_{[\gamma-\gamma^0_{\rm per}]}(y)}{|x-y|}dx\,dy+\frac12\int_{\mathbb{R}^3}\int_{\mathbb{R}^3}\frac{\rho_{[\gamma-\gamma^0_{\rm per}]}(x)\rho_{[\gamma-\gamma^0_{\rm per}]}(y)}{|x-y|}dx\,dy.&\qquad(5.19)\end{aligned}$$

The two terms in the left-hand side of (5.19) involve the usual reduced HF energy in the whole space introduced in Chapter 1. They are not well-defined because $\mu_{\rm per}$ is periodic and because γ and $\gamma^0_{\rm per}$ have infinite ranks, but we shall be able to give a mathematical meaning to the right-hand side, exploiting the fact that $Q:=\gamma-\gamma^0_{\rm per}$ induces a small perturbation of the reference state $\gamma^0_{\rm per}$. The formal computation (5.19) will be justified by means of thermodynamic limit arguments.

Similarly as in Chapter 4, we have to properly define our energy functional. As usual, we use the shorthand notation

$$Q^{--}:=\gamma^0_{\rm per}Q\gamma^0_{\rm per},\quad Q^{++}:=(1-\gamma^0_{\rm per})Q(1-\gamma^0_{\rm per}).$$

We introduce the Banach space

$$\mathcal{Q}=\left\{Q^*=Q\in\mathfrak{S}_2,\ :\ |\nabla|Q\in\mathfrak{S}_2,\ Q^{++},Q^{--}\in\mathfrak{S}_1,\ |\nabla|Q^{++}|\nabla|\in\mathfrak{S}_1,\ |\nabla|Q^{--}|\nabla|\in\mathfrak{S}_1\right\},$$

endowed with its natural norm

$$\begin{aligned}\|Q\|_{\mathcal{Q}}:=\|Q\|_{\mathfrak{S}_2}+\left\|Q^{++}\right\|_{\mathfrak{S}_1}+\left\|Q^{--}\right\|_{\mathfrak{S}_1}&\\ +\left\||\nabla|Q\right\|_{\mathfrak{S}_2}+\left\||\nabla|Q^{++}|\nabla|\right\|_{\mathfrak{S}_1}+\left\||\nabla|Q^{--}|\nabla|\right\|_{\mathfrak{S}_1}.&\qquad(5.20)\end{aligned}$$

For such states we can define a generalized charge similarly to what we have done in the previous Chapter, by introducing

$$\mathrm{Tr}_{\gamma^0_{\rm per}}(Q) := \mathrm{Tr}(Q^{++} + Q^{--}).$$

Again we refer to Appendix A for general properties of this generalization of the trace functional. The convex set on which the energy will be defined is

$$\mathcal{K} := \big\{Q \in \mathcal{Q} \mid -\gamma^0_{\rm per} \leq Q \leq 1 - \gamma^0_{\rm per}\big\}. \tag{5.21}$$

In order to define properly the energy of Q, we need to associate a density ρ_Q with any state $Q \in \mathcal{K}$. We recall that $\mathcal{C}$ is the Coulomb space containing all functions f such that $D(f,f) < \infty$. The density ρ_Q was defined in [9] by means of a duality argument:

Proposition 5.1 (Definition of the density [9]). *Assume that **(A1)** holds true and let $Q \in \mathcal{Q}$. Then $QV \in \mathfrak{S}_1^{\gamma^0_{\rm per}}$ for any $V = V_1 + V_2 \in \mathcal{C}' + \big(L^2(\mathbb{R}^3) \cap L^\infty(\mathbb{R}^3)\big)$ and moreover there exists a constant C (independent of Q and V) such that*

$$|\mathrm{Tr}_{\gamma^0_{\rm per}}(QV)| \leq C\,\|Q\|_{\mathcal{Q}}\,(\|V_1\|_{\mathcal{C}'} + \|V_2\|_{L^2(\mathbb{R}^3)}).$$

Thus the linear form $V \in \mathcal{C}' + \big(L^2(\mathbb{R}^3) \cap L^\infty(\mathbb{R}^3)\big) \mapsto \mathrm{Tr}_{\gamma^0_{\rm per}}(QV)$ can be continuously extended to $\mathcal{C}' + L^2(\mathbb{R}^3)$ and there exists a uniquely defined function $\rho_Q \in \mathcal{C} \cap L^2(\mathbb{R}^3)$ such that

$$\forall V = V_1 + V_2 \in \mathcal{C}' + \big(L^2(\mathbb{R}^3) \cap L^\infty(\mathbb{R}^3)\big)\,, \quad \langle \rho_Q, V_1\rangle_{\mathcal{C},\mathcal{C}'} + \int_{\mathbb{R}^3} \rho_Q V_2 = \mathrm{Tr}_{\gamma^0_{\rm per}}(QV).$$

The linear map $Q \in \mathcal{Q} \mapsto \rho_Q \in \mathcal{C} \cap L^2(\mathbb{R}^3)$ is continuous:

$$\|\rho_Q\|_{\mathcal{C}} + \|\rho_Q\|_{L^2(\mathbb{R}^3)} \leq C\,\|Q\|_{\mathcal{Q}}\,.$$

Eventually when $Q \in \mathfrak{S}_1 \subset \mathfrak{S}_1^{\gamma^0_{\rm per}}$, then $\rho_Q(x) = Q(x,x)$ where $Q(x,y)$ is the integral kernel of Q.

Proposition 5.1 is similar to Lemma 4.1 in Chapter 4. Assuming that **(A1)** holds true, we are now in a position to give a rigorous sense to the right-hand side of (5.19) for $\gamma - \gamma^0_{\rm per} = Q \in \mathcal{K}$. In the sequel, we use the following *notation* for any $Q \in \mathcal{Q}$:

$$\mathrm{Tr}_{\gamma^0_{\rm per}}(H^0_{\rm per}Q) := \mathrm{Tr}(|H^0_{\rm per} - \kappa|^{1/2}(Q^{++} - Q^{--})|H^0_{\rm per} - \kappa|^{1/2}) + \kappa\mathrm{Tr}_{\gamma^0_{\rm per}}(Q) \tag{5.22}$$

where κ is an arbitrary real number in the gap $(\Sigma^+_Z, \Sigma^-_{Z+1})$ (this expression can be proved to be independent of κ, see [9]). Then we define the energy of any state $Q \in \mathcal{K}$ as

$$\boxed{\mathcal{E}^\nu(Q) := \mathrm{Tr}_{\gamma^0_{\rm per}}(H^0_{\rm per}Q) - D(\rho_Q,\nu) + \frac{1}{2}D(\rho_Q,\rho_Q).} \tag{5.23}$$

It was shown in [9] that the energy is well-defined for all $Q \in \mathcal{K}$.

5.2.2 Existence of minimizers

The next step is to study the existence and the properties of minimizers. Like in the previous chapter, we will minimize while imposing a charge constraint $\mathrm{Tr}_{\gamma^0_{\rm per}}(Q) = q$. Similarly, one can minimize the functional without any constraint but with a chemical potential. Indeed, arguing like in Chapter 4, one easily sees that

$$\forall Q \in \mathcal{K}, \qquad \mathcal{E}^\nu(Q) - \epsilon_{\rm F}\mathrm{Tr}_{\gamma^0_{\rm per}}(Q) \geq -\frac{1}{2}D(\nu,\nu).$$

Even when ν is a delta function, the energy can be shown to be bounded from below as was explained in [12]. Concerning the existence of minimizers, the following was proved in [9] (the extension to a pointwise defect is taken care of in [12] but we do not state an explicit result).

Theorem 5.4 (Existence of minimizers with a chemical potential [9])**.** *Let* $\nu \in L^1(\mathbb{R}^3) \cap L^2(\mathbb{R}^3)$*,* $Z \in \mathbb{N} \setminus \{0\}$ *and assume that* **(A1)** *holds. Then for any* $\epsilon_F \in (\Sigma_Z^+, \Sigma_{Z+1}^-)$*, there exists a minimizer* $\bar{Q} \in \mathcal{K}$ *for the following variational problem:*

$$\boxed{E^\nu_{\epsilon_F} := \inf\{\mathcal{E}^\nu(Q) - \epsilon_F \mathrm{Tr}_{\gamma^0_{\rm per}}(Q),\ Q \in \mathcal{K}\} > -\infty}. \tag{5.24}$$

Problem (5.24) *may have several minimizers, but they all share the same density* $\bar\rho = \rho_{\bar{Q}}$*. Any minimizer* $\bar{Q}$ *of* (5.24) *satisfies the self-consistent equation*

$$\left\{\begin{array}{l} \bar{Q} = \chi_{(-\infty,\epsilon_F)}(H_{\bar{Q}}) - \gamma^0_{\rm per} + \delta, \\ H_{\bar{Q}} = H^0_{\rm per} + (\rho_{\bar{Q}} - \nu) * |\cdot|^{-1} \end{array}\right. \tag{5.25}$$

where δ *is a finite rank self-adjoint operator satisfying* $0 \leq \delta \leq 1$ *and* $\mathrm{Ran}(\delta) \subseteq \ker(H_{\bar{Q}} - \epsilon_F)$*.*

It is easily seen that $(\rho_{\bar{Q}} - \nu) * |\cdot|^{-1}$ is a compact perturbation of $H^0_{\rm per}$, implying that $H_{\bar{Q}}$ is self-adjoint on $\mathcal{D}(H^0_{\rm per}) = \mathcal{D}(-\Delta) = H^2(\mathbb{R}^3)$ and that $\sigma_{\rm ess}(H_{\bar{Q}}) = \sigma(H^0_{\rm per})$. Thus the discrete spectrum of $H_{\bar{Q}}$ is composed of isolated eigenvalues of finite multiplicity, possibly accumulating at the ends of the bands, see Fig. 5.2.

Recall that the charge of the minimizing state $\bar{Q}$ obtained in Theorem 5.4 is defined as $\mathrm{Tr}_{\gamma^0_{\rm per}}(\bar{Q})$. Similarly to what was explained in Chapter 4, it can be proved by perturbation theory that for any fixed ϵ_F, there exists a constant $C(\epsilon_F)$ such that when $D(\nu,\nu) \leq C(\epsilon_F)$, one has $\ker(H_{\bar{Q}} - \epsilon_F) = \{0\}$ and $\mathrm{Tr}_{\gamma^0_{\rm per}}(\bar{Q}) = 0$, i.e. the minimizer of the energy with chemical potential ϵ_F is a neutral perturbation of the periodic Fermi sea.

The proof of Theorem 5.4 uses the same general ideas as in the relativistic case. There are however several complications, the biggest one being that there is no cut-off in Fourier space.

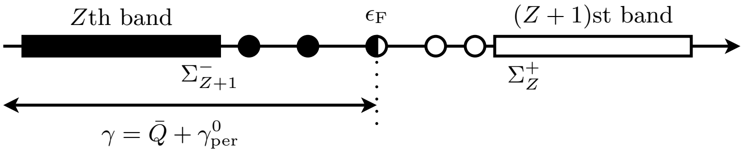


Figure 5.2: Spectrum of $H_{\bar{Q}}$.

We have stated the existence of minimizers for any chemical potential in the gap of the periodic operator $H^0_{\rm per}$, but of course the total charge $\mathrm{Tr}_{\gamma^0_{\rm per}}(\bar{Q})$ of the obtained solution was unknown *a priori.* We now tackle the more subtle problem of minimizing the energy while *imposing a charge constraint.* Mathematically this is more difficult because although the energy $\mathcal{E}^\nu(Q)$ is convex on $\mathcal{K}$ and weakly lower semi-continuous (wlsc) for the weak-$*$ topology of $\mathcal{Q}$, the $\gamma^0_{\rm per}$-trace functional $Q \in \mathcal{K} \mapsto \mathrm{Tr}_{\gamma^0_{\rm per}}(Q)$ is continuous but *not* wlsc for the weak-$*$ topology of $\mathcal{Q}$: in principle it is possible that a (positive or negative) part of the charge of a minimizing sequence for the charge-constrained minimization problem escapes to infinity, leaving at the limit a state of a different (lower or higher) charge. In fact, we can prove that a minimizer exists under a charge constraint, if and only if some binding conditions hold, the role of which being to prevent the lack of compactness.

We introduce the minimization problem with a real charge constraint $q \in \mathbb{R}$:

$$\boxed{E^\nu(q) := \inf\{\mathcal{E}^\nu(Q),\ Q \in \mathcal{K},\ \mathrm{Tr}_{\gamma^0_{\rm per}}(Q) = q\}.} \tag{5.26}$$

When no defect is present, $E^0(q)$ can be computed explicitly [9]:

$$E^0(q) = \left\{\begin{array}{ll} \Sigma^-_{Z+1} q & \text{when } q \geq 0 \\ \Sigma^+_Z q & \text{when } q \leq 0. \end{array}\right.$$

The following result, similar to Theorem 4.11 in Chapter 4, was proved in [9]:

Theorem 5.5 (Existence of minimizers under a charge constraint [9]). *Let $\nu \in L^1(\mathbb{R}^3) \cap L^2(\mathbb{R}^3)$, $Z \in \mathbb{N} \setminus \{0\}$ and assume that* **(A1)** *holds. The following assertions are equivalent:*

(a) *Problem* (5.26) *admits a minimizer $\bar{Q}$;*

(b) *Every minimizing sequence for* (5.26) *is precompact in $\mathcal{Q}$ and converges towards a minimizer $\bar{Q}$ of* (5.26)*;*

(c) $\forall q' \in \mathbb{R} \setminus \{0\}, \quad E^\nu(q) < E^\nu(q - q') + E^0(q')$.

Assume that the equivalent conditions **(a)**, **(b)** *and* **(c)** *above are fulfilled. In this case, the minimizer $\bar{Q}$ is not necessarily unique, but all the minimizers share the same density $\bar{\rho} = \rho_{\bar{Q}}$. Besides, there exists $\epsilon_F \in [\Sigma_Z^+, \Sigma_{Z+1}^-]$ such that the obtained minimizer $\bar{Q}$ is a global minimizer for $E_{\epsilon_F}^\nu$ defined in* (5.24). *It solves Equation* (5.25) *for some $0 \leq \delta \leq 1$ with $\mathrm{Ran}(\delta) \subseteq \ker(H_{\bar{Q}} - \epsilon_F)$. The operator δ is finite rank if $\epsilon_F \in (\Sigma_Z^+, \Sigma_{Z+1}^-)$ and trace-class if $\epsilon_F \in \{\Sigma_Z^+, \Sigma_{Z+1}^-\}$.*

Additionally the set of q's in $\mathbb{R}$ satisfying the above equivalent conditions is a non-empty closed interval $I \subseteq \mathbb{R}$. This is the largest interval on which $q \mapsto E^\nu(q)$ is strictly convex.

Unfortunately, and contrarily to the relativistic case, we have not been able to verify the binding condition **(c)**. In the QED case, we first proved that ρ_Q must be L^1 and use this fact to show that neutral atoms always have a minimizer. Here it is not expected that ρ is in L^1 (this will be explained below in Section 5.2.4) and proving **(c)** would probably require a better understanding of the properties of the density ρ_Q for a minimizer, when it exists.

5.2.3 Thermodynamic limit of the supercell model

Similarly as before, we now derive the reduced Hartree-Fock model with defect that we have just introduced as the thermodynamic limit of the super-cell method. First we need to periodize the defect ν with respect to the large box Λ_L, for instance by defining

$$\nu_L(x) := \sum_{z \in \mathbb{Z}^3} (\mathbb{1}_{\Lambda_L} \nu)(\cdot - Lz).$$

The reduced Hartree-Fock energy functional of the supercell model with defect is then defined for $\gamma \in \mathcal{P}_{\mathrm{sc},L}$ as

$$\boxed{\mathcal{E}_{\mathrm{sc},L}^\nu(\gamma) = \mathrm{Tr}_{L^2_{\mathrm{per}}(\Lambda_L)} \left(-\frac{1}{2}\Delta\gamma \right) + \frac{1}{2} D_{G_L} \left(\rho_\gamma - \mu_{\mathrm{per}} - \nu_L, \rho_\gamma - \mu_{\mathrm{per}} - \nu_L \right).}$$

For $\epsilon_F \in (\Sigma_Z^+, \Sigma_{Z+1}^-)$, we consider the following minimization problem

$$\boxed{I_{\mathrm{sc},L,\epsilon_F}^\nu = \inf \left\{ \mathcal{E}_{\mathrm{sc},L}^\nu(\gamma) - \epsilon_F \mathrm{Tr}_{L^2_{\mathrm{per}}(\Lambda_L)}(\gamma), \ \gamma \in \mathcal{P}_{\mathrm{sc},L} \right\}.} \tag{5.27}$$

Theorem 5.6 (Thermodynamic limit of the supercell model with defect [9]). *Let $Z \in \mathbb{N} \setminus \{0\}$. Assume that* **(A1)** *holds and fix some $\epsilon_F \in (\Sigma_Z^+, \Sigma_{Z+1}^-)$. Then one has*

$$\lim_{L \to \infty} \left(I_{\mathrm{sc},L,\epsilon_F}^\nu - I_{\mathrm{sc},L,\epsilon_F}^0 \right) = E_{\epsilon_F}^\nu - \int_{\mathbb{R}^3} \nu \left((\rho_{\gamma_{\mathrm{per}}^0} - \mu_{\mathrm{per}}) *_{\mathcal{R}} G \right) + \frac{1}{2} D(\nu, \nu). \tag{5.28}$$

Additionally, if $\gamma_{\mathrm{sc},L}^\nu$ denotes a minimizer for (5.27)*, then one has, up to extraction of a subsequence,*

$$(\gamma_{\mathrm{sc},L}^\nu - \gamma_{\mathrm{sc},L}^0)(x, y) \to \bar{Q}(x, y)$$

weakly in $H^1_{\mathrm{loc}}(\mathbb{R}^3 \times \mathbb{R}^3)$ and strongly in $L^2_{\mathrm{loc}}(\mathbb{R}^3 \times \mathbb{R}^3)$, where $\bar{Q}$ is a minimizer of (5.24)*, as obtained in Theorem 5.4. Besides,*

$$\rho_{\gamma_{\mathrm{sc},L}^\nu} - \rho_{\gamma_{\mathrm{sc},L}^0} \to \bar{\rho}$$

weakly in $L^2_{\mathrm{loc}}(\mathbb{R}^3)$, where $\bar{\rho}$ is the common density of all the minimizers of (5.24).

In numerical simulations, the right-hand side of (5.28) is approximated by $I^{\nu}_{\mathrm{sc},L,\epsilon_F} - I^{0}_{\mathrm{sc},L,\epsilon_F}$ for a given value of L. This approach has several drawbacks. First, the values of L that lead to tractable numerical simulations are in many cases much too small to obtain a correct estimation of the limit $L \to \infty$. Second, it is not easy to extend this method for computing $E^{\nu}_{\epsilon_F}$, to the direct evaluation of $E^{\nu}(q)$ for a given q (i.e. the energy of a defect with a prescribed total charge). The formalism introduced in this chapter suggests an alternative way for computing energies of defects in crystalline materials that will be detailed in Section 5.3.

5.2.4 Properties of minimizers: the dielectric permittivity of a crystal

In this subsection, we review a recent work [12] with É. Cancès in which we have investigated the regularity properties of minimizers Q of our rHF functional for crystals with defects. In this section, we will not assume anymore that the lattice is necessarily $\mathcal{R} = \mathbb{Z}^3$. Similarly we do not assume *a priori* that the set of nuclei has any special symmetry (except, of course that it must be $\mathcal{R}$-periodic). As we have already mentioned, all the previous results are valid in this general setting. The reason of this sudden change is that, as we will see, isotropic crystals are a bit special when dealing with properties of minimizers.

In the previous chapter on no-photon QED, a crucial role was played by the first order density, i.e. by the linear response of the model to a small external density ν. This in particular led to the concept of charge renormalization via the properties of the function $B_\Lambda(k)$ at zero. In the crystal case the same kind of behavior will occur, with however some important differences. First there is no cut-off in the rHF model of the crystal, and there will not be any renormalization issue. Like in the previous chapter, minimizers will not be trace-class and the Fermi sea will partly screen the defect (by a finite amount of charge). However we will see that, contrarily to the Dirac case, in most cases the density is *not even* L^1 and it is believed to have a peculiar behavior at zero in Fourier space. Hence it is not really possible to define an observed charge. This makes the study of the regularity of ρ_Q (hence the proof of the existence of neutral systems) more difficult than in the reduced BDF model.

5.2.4.1 The linear response

The first order density will be described by the following map

$$\mathcal{L}(f) := -\rho\left[\frac{1}{2i\pi}\oint_{\mathscr{C}}\left(z - H^0_{\mathrm{per}}\right)^{-1} f * \frac{1}{|\cdot|}\left(z - H^0_{\mathrm{per}}\right)^{-1} dz\right], \tag{5.29}$$

where $\mathscr{C}$ is a smooth curve in the complex plane enclosing the whole spectrum of H^0_{per} below ϵ_{F}, crossing the real line at ϵ_{F} and at some $c < \inf\sigma(H^0_{\mathrm{per}})$ (the above quantity is independent of $\mathscr{C}$). The reason why we have put a minus sign in (5.29) is very simple: in the rHF nonlinear case, we will have

$$\rho_Q = \mathcal{L}(\nu - \rho_Q) + \tilde{r}_2$$

where $\tilde{r}_2$ contains the higher order terms, and which we will rewrite as

$$(1 + \mathcal{L})(\nu - \rho_Q) = \nu - \tilde{r}_2. \tag{5.30}$$

This motivates the following result of [12], which is central in the mathematical analysis of the dielectric response of crystals.

Theorem 5.7 (Properties of $\mathcal{L}(\rho)$ when $\rho \in \mathcal{C}$ or when $\rho \in L^1$, [12]).

1. *$\mathcal{L}$ defines a bounded nonnegative self-adjoint operator on $\mathcal{C}$. Hence $1 + \mathcal{L}$, considered as an operator on $\mathcal{C}$, is invertible and bicontinuous from $\mathcal{C}$ to $\mathcal{C}$.*

2. Let $\rho \in L^1(\mathbb{R}^3)$*. Then,* $\mathcal{L}(\rho) \in L^2(\mathbb{R}^3) \cap \mathcal{C}$*,* $\widehat{\mathcal{L}(\rho)}$ *is continuous on* $\Gamma^* \setminus \{0\}$*, and for all* $\sigma \in S^2$ *(the unit sphere of* $\mathbb{R}^3$*),*

$$\lim_{\eta \to 0^+} \widehat{\mathcal{L}(\rho)}(\eta\sigma) = (\sigma^T L \sigma)\,\widehat{\rho}(0) \tag{5.31}$$

with L *being the* 3×3 *matrix defined by*

$$k^T L k = \frac{8\pi}{|\Gamma|} \sum_{n=1}^{N} \sum_{n'=N+1}^{+\infty} \fint_{\Gamma^*} \frac{\left|\langle (k \cdot \nabla_x) u_{n,q}, u_{n',q} \rangle_{L^2_{\rm per}(\Gamma)}\right|^2}{\left(\epsilon_{n',q} - \epsilon_{n,q}\right)^3}\, dq, \tag{5.32}$$

where the $\epsilon_{n,q}$*'s and the* $u_{n,q}$*'s are the eigenvalues and the eigenvectors arising in the Bloch-Floquet spectral decomposition of* $(H^0_{\rm per})_q : L^2_{\rm per}(\Gamma) \to L^2_{\rm per}(\Gamma)$*:*

$$(H^0_{\rm per})_q = \sum_{n=1}^{\infty} \epsilon_{n,q} |u_{n,q}\rangle\langle u_{n,q}|. \tag{5.33}$$

Additionally, $L \geq 0$ *and*

$$L_0 = \frac{1}{3}{\rm Tr}(L) > 0. \tag{5.34}$$

Note that (5.34) implies that $\sigma \mapsto \sigma^T L \sigma$ does not vanish for at least a set of nonzero measure. Theorem 5.7 shows that $\mathcal{L}(\rho)$ is not in general a function of $L^1(\mathbb{R}^3)$ even when $\rho \in L^1(\mathbb{R}^3)$, as when L is not constant, $\widehat{\mathcal{L}(\rho)}$ is not continuous at zero (note that $L \equiv L_0$ characterizes isotropic dielectric materials).

5.2.4.2 Properties of minimizers

Let us now come back to the reduced Hartree-Fock framework and the decay properties of minimizers. The following is taken from [12]:

Theorem 5.8 (Properties of the nonlinear rHF ground state for perturbed crystals [12])**.** *Let* $\epsilon_{\rm F}$ *in the gap* $(\Sigma^+_Z, \Sigma^-_{Z+1})$*. Let* $\nu \in L^1(\mathbb{R}^3) \cap L^2(\mathbb{R}^3)$ *be such that* $\int_{\mathbb{R}^3} \nu \neq 0$ *and* $\|\nu * |\cdot|^{-1}\|_{L^2 + \mathcal{C}'}$ *is small enough. Then the variational problem* (5.24) *has a unique minimizer* $Q_{\nu,\epsilon_{\rm F}}$*. It satisfies* ${\rm Tr}_0(Q_{\nu,\epsilon_{\rm F}}) = 0$ *but it is* <u>*not*</u> *trace-class. If additionally the matrix* L *defined in Theorem 5.7 is not proportional to the identity, then* $\rho_{\nu,\epsilon_{\rm F}}$ *is* <u>*not*</u> *in* $L^1(\mathbb{R}^3)$*.*

The proof of Theorem 5.8 is a simple consequence of Theorem 5.7 and of the continuity properties of higher order terms for an L^1 density ρ.

As previously mentioned, the situation $L \equiv L_0$ characterizes isotropic dielectric materials; it occurs in particular when $\mathcal{R}$ is a cubic lattice and $\rho^{\rm nuc}_{\rm per}$ has the symmetry of the cube. For anisotropic dielectric materials, $\sigma \mapsto \sigma^T L \sigma$ is not a constant function, so that $\rho_{\nu,\epsilon_{\rm F}} \notin L^1(\mathbb{R}^3)$.

Formula (5.34) for L is well-known in the Physics literature [2, 37]. However to our knowledge it was never mentioned that the fact that $L \neq 0$ is linked to the odd mathematical property that the operator $Q_{\nu,\epsilon_{\rm F}}$ is not trace-class when $\int_{\mathbb{R}^3} \nu \neq 0$.

For isotropic dielectric materials, $L = L_0$, and we conjecture that the density $\rho_{\nu,\epsilon_{\rm F}}$ is in $L^1(\mathbb{R}^3)$. In this case, one can define the total charge of the defect (including the self-consistent polarization of the Fermi sea) as $\int(\nu - \rho_{\nu,\epsilon_{\rm F}})$. For ν small enough, the Fermi sea formally stays neutral, ${\rm Tr}_0(Q_{\nu,\epsilon_{\rm F}}) = 0$, but it nevertheless screens partially the charge defect in such a way that the total observed charge gets multiplied by a factor $(1 + L_0)^{-1} < 1$:

$$\int_{\mathbb{R}^3} \left(\nu - \rho_{\nu,\epsilon_{\rm F}}\right) = \frac{\int_{\mathbb{R}^3} \nu}{1 + L_0},$$

similarly to (4.54) in Chapter 4. Contrarily to the Fermi sea of periodic crystals, the rBDF free vacuum of Chapter 4 is not only isotropic but also homogeneous, and the mathematical analysis can be pushed further. Extending these results to the case of isotropic crystals seems a very challenging task.

When L is not proportional to the identity (anisotropic dielectric crystals), it is not possible to define the observed charge of the defect as the integral of $\nu - \rho_{\nu,\epsilon_{\rm F}}$ since $\rho_{\nu,\epsilon_{\rm F}}$ is not an integrable function. Understanding the regularity properties of the Fourier transform of $\rho_{\nu,\epsilon_{\rm F}}$ is then a very interesting problem. In the next section, we consider a certain limit related to homogenization in which only the first order term plays a role and for which one can analyse the limit in details.

5.2.4.3 Macroscopic dielectric permittivity

In this section, we focus on the electrostatic potential

$$V = (\nu - \rho_{\nu,\epsilon_{\rm F}}) * |\cdot|^{-1} \tag{5.35}$$

generated by the total charge of the defect and we study it in a certain limit.

We note that the self-consistent equation (5.30) can be rewritten as

$$\nu - \rho_{\nu,\epsilon_{\rm F}} = (1+\mathcal{L})^{-1}\nu - (1+\mathcal{L})^{-1}\tilde{r}_2. \tag{5.36}$$

Therefore for the nonlinear rHF model, the linear response at the level of the density is given by the operator $(1+\mathcal{L})^{-1}$. Even when $\nu \in L^1(\mathbb{R}^3)$, applying the operator $(1+\mathcal{L})^{-1}$ creates some discontinuities in the Fourier domain for the corresponding first order term $(1+\mathcal{L})^{-1}\nu$ in Equation (5.36). If we knew that the higher order term $\tilde{r}_2$ is better behaved, it would be possible to deduce the exact regularity of $\widehat{\rho}_{\nu,\epsilon_{\rm F}}$. We will now consider a certain limit of (5.36) by means of a homogenization argument, for which the second order term disappears. This will give an illustration of the expected properties of the density in Fourier space at the origin. For this purpose, we fix some $\nu \in L^1(\mathbb{R}^3) \cap L^2(\mathbb{R}^3)$ and introduce for all $\eta > 0$ the rescaled density

$$\nu_\eta(x) := \eta^3 \nu(\eta x).$$

We then denote by V_ν^η the total potential generated by ν_η, i.e.

$$V_\nu^\eta := (\nu - \rho_{\nu_\eta,\epsilon_{\rm F}}) * \frac{1}{|\cdot|}, \tag{5.37}$$

and define the rescaled potential

$$W_\nu^\eta(x) := \eta^{-1}\, V_\nu^\eta\left(\eta^{-1}x\right). \tag{5.38}$$

Note that the scaling parameters have been chosen in such a way that in the absence of dielectric response (i.e. for $\mathcal{L} = 0$, $\tilde{r}_2 = 0$), one has $W_\nu^\eta = \nu * |\cdot|^{-1}$ for all $\eta > 0$.

Theorem 5.9 (Macroscopic Dielectric Permittivity [12]). *There exists a 3×3 symmetric matrix $\epsilon_{\rm M} > 1$ such that for all $\nu \in L^1(\mathbb{R}^3)\cap L^2(\mathbb{R}^3)$, the rescaled potential W_ν^η defined by (5.38) converges to W_ν weakly in $\mathcal{C}'$ when η goes to zero, where W_ν is the unique solution in $\mathcal{C}'$ to the elliptic equation*

$$\boxed{-\mathrm{div}(\epsilon_{\rm M}\nabla W_\nu) = 4\pi\nu.}$$

The matrix $\epsilon_{\rm M}$ is proportional to the identity matrix if the crystal has the symmetry of the cube.

From a physical viewpoint, the matrix $\epsilon_{\rm M}$ is the electronic contribution to the macroscopic dielectric tensor of the host crystal. Note the other contribution, originating from the displacements of the nuclei [29], is not taken into account here.

The matrix $\epsilon_{\rm M}$ can be computed from the Bloch-Floquet decomposition of $H^0_{\rm per}$ as follows. We consider the operator $\widetilde{\epsilon} = v_{\rm c}^{1/2}(1+\mathcal{L})v_{\rm c}^{-1/2}$ where $v_{\rm c}(\nu) := \nu * |\cdot|^{-1}$ is the Coulomb operator. By the properties of $\mathcal{L}$, $\widetilde{\epsilon}$ is a bounded self-adjoint operator on $L^2(\mathbb{R}^3)$ such that $\widetilde{\epsilon} \geq 1$. As $\widetilde{\epsilon}$ commutes with the translations of the lattice, it can be represented by the Bloch matrices $([\widetilde{\epsilon}_{KK'}(q)]_{K,K'\in\mathcal{R}^*})_{q\in\Gamma^*}$ which means

$$\forall f \in L^2(\mathbb{R}^3), \quad \widehat{\widetilde{\epsilon}f}(q+K) = \sum_{K'\in\mathcal{R}^*} \widetilde{\epsilon}_{KK'}(q)\widehat{f}(q+K')$$

for almost all $q \in \Gamma^*$ and $K \in \mathcal{R}^*$. Similarly we may define the Bloch matrix $\widetilde{\epsilon}^{-1}_{KK'}$ of $\widetilde{\epsilon}^{-1} = v_{\rm c}^{1/2}(1+\mathcal{L})^{-1}v_{\rm c}^{-1/2}$. It is shown in [12] that $\tilde{\epsilon}_{K,K'}(\eta\sigma)$ has a limit when η goes to 0^+ for all fixed $\sigma \in S^2$. Indeed one has $\tilde{\epsilon}_{0,0}(\eta\sigma) \to 1 + \sigma^T L\sigma$ where L is the 3×3 non-negative symmetric matrix defined in (5.32). When $K, K' \neq 0$, $\tilde{\epsilon}_{K,K'}(\eta\sigma)$ has a limit for $\eta \to 0$, which is independent of σ and which we simply denote as $\tilde{\epsilon}_{K,K'}(0)$. When $K = 0$ but $K' \neq 0$, the limit is a linear function of σ: for all $K' \in \mathcal{R}^*\setminus\{0\}$, $\tilde{\epsilon}_{0,K'}(\eta\sigma) \to \beta_{K'}\cdot\sigma$, for some $\beta_{K'} \in \mathbb{C}^3$. The electronic contribution to the macroscopic dielectric permittivity is the 3×3 symmetric tensor defined as [6]

$$\forall k \in \mathbb{R}^3, \quad k^T\epsilon_{\rm M}k = \lim_{\eta\to 0^+} \frac{|k|^2}{[\widetilde{\epsilon}^{-1}]_{00}(\eta k)}. \tag{5.39}$$

By the Schur complement formula, one has

$$\frac{1}{[\widetilde{\epsilon}^{-1}]_{00}(\eta k)} = \tilde{\epsilon}_{00}(\eta k) - \sum_{K,K'\neq 0} \tilde{\epsilon}_{0,K}(\eta k)[C(\eta k)^{-1}]_{K,K'}\tilde{\epsilon}_{K',0}(\eta k)$$

where $C(\eta k)^{-1}$ is the inverse of the matrix $C(\eta k) = [\tilde{\epsilon}_{KK'}(\eta k)]_{K,K'\in\mathcal{R}^*\setminus\{0\}}$. This leads to

$$\boxed{\epsilon_{\rm M} = 1 + L - \sum_{K,K'\in\mathcal{R}^*\setminus\{0\}} \beta_K[C(0)^{-1}]_{K,K'}\beta^*_{K'}.} \tag{5.40}$$

Formula (5.40) has been used in numerical simulations for estimating the macroscopic dielectric permittivity of real insulators and semiconductors [6, 19, 20, 15, 16]. Direct methods for evaluating $\epsilon_{\rm M}$, bypassing the inversion of the matrix $C(0)$, have also been proposed [32, 23].

5.3 Variational approximation and numerical results in 1D

Let us now come to the discretization of problem (5.26), as was presented in [10]. If one discretizes (5.26) in a local basis without taking care of the constraint $Q \in \mathcal{K}$, there is a risk to obtain meaningless numerical results, due to a spectral pollution phenomenon, as will be explained in Chapter 6. On the other hand, selecting a basis set which respects the decomposition

$$L^2(\mathbb{R}^3) = \mathfrak{H}_- \oplus \mathfrak{H}_+ = \gamma^0_{\rm per}L^2(\mathbb{R}^3) \oplus (\gamma^0_{\rm per})^\perp L^2(\mathbb{R}^3)$$

will lead to a well-behaved variational approximation of (5.26), see Section 6.2.3.1 in Chapter 6.

Let $V^h_\pm$ be finite-dimensional subspaces of the occupied and virtual spaces $\mathfrak{H}_\pm$ of the reference perfect crystal. Consider the finite-dimensional subspace $V^h = V^h_- \oplus V^h_+$ of $L^2(\mathbb{R}^3)$. Let $(\varphi_1, \cdots, \varphi_{m_-})$ (resp. $(\varphi_{m_-+1}, \cdots, \varphi_{N_b})$) be an orthonormal basis of V^h_- (resp. of V^h_+). We denote for simplicity $m_+ := N_b - m_-$. The approximation set for Q consists of the finite-rank operators

$$Q = \sum_{i,j=1}^{N_b} Q^h_{ij}|\varphi_i\rangle\langle\varphi_j| \tag{5.41}$$

with $Q^h \in \mathcal{K}^h = \left\{Q^h = [Q^h]^*,\ 0 \leq \mathcal{I} + Q^h \leq 1\right\}$, where $\mathcal{I}$ is the $N_b \times N_b$ block diagonal matrix

$$\mathcal{I} = \left[\begin{array}{cc} 1_{m_-} & 0 \\ 0 & 0_{m_+} \end{array}\right].$$

Minimizing over Q's of this form, one gets an *upper bound* to the true energy $E^\nu(q)$. As $Q^h \in \mathcal{K}^h$ with $\mathrm{Tr}(Q^h) = q$ if and only if

$$\mathcal{I} + Q^h \in \left\{D = D^T \in \mathbb{R}^{2N_b},\, D^2 \leq D,\, \mathrm{Tr}(D) = q + m_-\right\},$$

the corresponding finite-dimensional problem can be solved using relaxed constrained algorithms [11, 8, 22].

The question is now to build spaces V_-^h and V_+^h that provide good approximations to (5.26). A natural choice proposed and tested in [10] is to use Wannier functions [36] of the reference perfect crystal. Wannier functions $\{w_k\}$ are in general defined in such a way that w_k belongs to the spectral subspace associated with the kth band and $\{w_k(\cdot - a)\}_{a\in\mathbb{Z}^3}$ forms a basis of this spectral subspace. One can take

$$w_k(x) = \int_{\Gamma^*} u_k(\xi, x) d\xi \tag{5.42}$$

where $u_k(\xi,\cdot) \in L^2_\xi$ is for any $\xi \in \Gamma^*$ an eigenvector of $(H^0_{\mathrm{per}})_\xi$ corresponding to the kth eigenvalue $\epsilon_k(\xi)$. The so-defined $\{w_k(\cdot - a)\}_{a\in\mathbb{Z}^3}$ are mutually orthogonal. Formula (5.42) does not define w_k uniquely since the $u_k(\xi,x)$ are in the best case only known up to a phase. Choosing the right phase, one can prove that when the kth band is isolated from other bands, w_k decays exponentially [27].

More generally, instead of using only one band (i.e. one eigenfunction $u_k(\xi,x)$), one can use K different bands for which it is possible to construct K exponentially localized Wannier functions as soon as the union of the K bands is isolated from the rest of the spectrum [28, 7]. The union of the K bands is called a *composite band*. In our case, under assumption **(A1)**, we typically have a natural composite band corresponding to the first Z bands, and another one corresponding to the other bands (the latter is not bounded above).

We emphasize that the Wannier basis *does not* depend on the defect, and can be precalculated once and for all for a given perfect crystal. Another huge advantage is that since w_k decays fast, it will be localized over a certain number of unit cells of $\mathbb{Z}^3$. When there is a localized defect in the lattice (let's say at z), keeping only the Wannier functions $w_k(\cdot - a)$ with $a \in \mathbb{Z}^3 \cap B(z,R)$ for some radius $R > 0$ should already yield a very good approximation. This approximation can be improved by enlarging progressively the radius R.

Of course in practice exponentially localized Wannier functions are not simple to calculate. Marzari and Vanderbilt have defined the concept of *maximally localized Wannier functions* [26] which also form a basis of the associated composite bands, but are not necessarily orthogonal with each other. Several efficient methods are known to find these functions numerically.

To construct V_-^h, one can therefore select the maximally localized (generalized) Wannier functions of the occupied bands, that overlap with some ball $B(z,R)$ of radius R centered on the nuclear charge defect. To obtain a basis set for V_+^h, one can select a number of active (unoccupied) bands using an energy cut-off and retain the maximally localized (generalized) Wannier functions of the active bands that overlap with the same ball $B(z,R)$. The so-obtained basis set of the virtual space can be enriched by adding projected atomic orbitals of the atoms and ghost atoms involved in ν (using the localized Wannier functions of the occupied bands to project out the $\mathcal{H}_-$ component of atomic orbitals preserves the locality of these orbitals).

In order to illustrate the efficiency of the variational approximation presented above, the example of a one-dimensional (1D) model with Yukawa interaction potential was tackled in [10]. The energy functional reads

$$E_{1\mathrm{D}}(\gamma) = \mathrm{Tr}\left(-\frac{1}{2}\frac{d^2\gamma}{dx^2}\right) - D_\kappa(\rho_{\mathrm{nuc}}, \rho_\gamma) + \frac{1}{2} D_\kappa(\rho_\gamma, \rho_\gamma)$$

with

$$D_\kappa(f,g) = (A/2\kappa) \int_{\mathbb{R}} \int_{\mathbb{R}} f(x)\, e^{-\kappa\,|x-x'|}\, g(x')\, dx\, dx'.$$

In the numerical examples reported below, the host crystal is $\mathbb{Z}$-periodic and the nuclear density is a Dirac comb, i.e. $\rho_{\rm nuc} = Z \sum_{j\in\mathbb{Z}} \delta_j$, with Z a positive integer. The values of the parameters ($A = 10$ and $\kappa = 5$) have been chosen in such a way that the ground state kinetic and potential energies are of the same order of magnitude. The nuclear local defect is taken of the form

$$\nu = (Z-1)\delta_{0.25} - Z\delta_0.$$

This corresponds to moving one nucleus and lowering its charge by one unit.

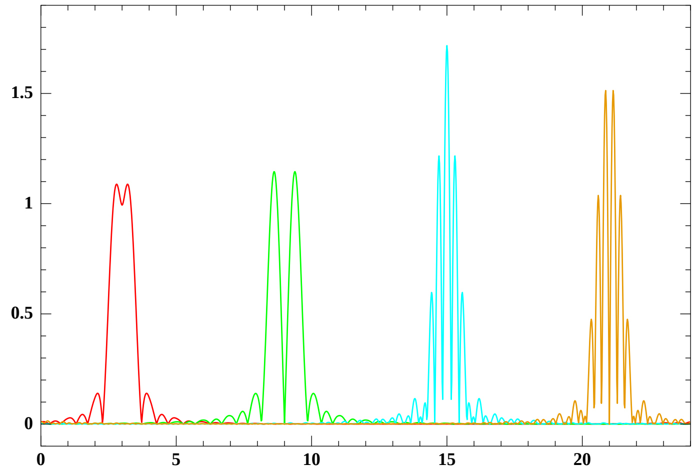


Figure 5.3: Modulus of MLWFs associated with the two occupied bands (left) and with the lowest two virtual bands (right).

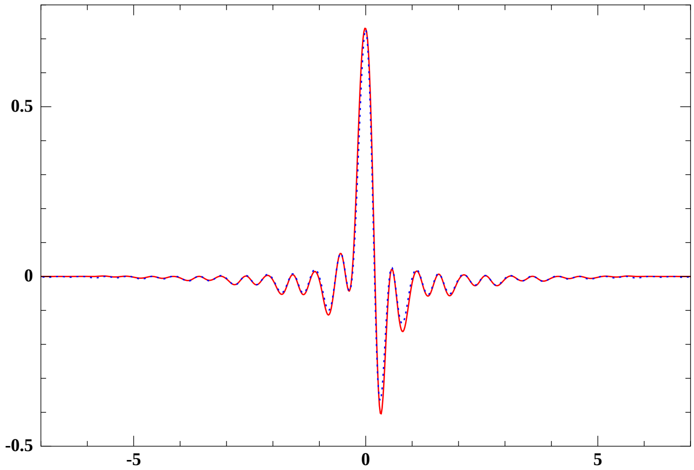


Figure 5.4: Density ρ_{Q^h} obtained with 28 MLWFs (line in red). The reference is a supercell calculation in a basis set of size 1224 (dashed line in blue).

The first stage of the calculation consists in solving the cell problem. For simplicity, we used a uniform discretization of the Brillouin zone $(-\pi, \pi]$, and a plane wave expansion of the crystalline orbitals.

The second stage is the construction of MLWFs. For this purpose, we make use of an argument specific to the one-dimensional case [33]: the MLWFs associated with the spectral projector γ are the eigenfunctions of the operator $\gamma x \gamma$. One first constructs N_e mother MLWFs (taking $\gamma = \gamma^0_{\rm per}$), then N_a mother MLWFs corresponding to the lowest N_a virtual bands (taking for γ the spectral projector associated with the lowest N_a virtual bands). The so-obtained mother MLWFs are represented on Fig. 5.3.

The third stage consists in constructing a basis set $(\varphi_j)_{1\le j\le N_b}$ of MLWFs by selecting translations of mother MLWFs that are closest to the local defect, and in computing the first-order density matrix of the form (5.41) which satisfies the constraints and minimizes the energy. The profile of the density ρ_{Q^h} obtained with $Z = 2$, $N_e = 2$, $N_a = 2$ and $N_b = 28$ is displayed on Fig. 5.4. It is compared with a reference supercell calculation with 1224 plane wave basis functions. A fairly good agreement is obtained with very few MLWFs.

The implementation of our method in the Quantum Espresso suite of programs [1], in the 3D case is work in progress.

References

[1] *Quantum Espresso*. See `http://www.quantum-espresso.org/`.

[2] S. L. Adler, *Quantum theory of the dielectric constant in real solids*, Phys. Rev., 126 (1962), pp. 413–420.

[3] V. Bach, *Error bound for the Hartree-Fock energy of atoms and molecules*, Commun. Math. Phys., 147 (1992), pp. 527–548.

[4] V. Bach, E. H. Lieb, M. Loss, and J. P. Solovej, *There are no unfilled shells in unrestricted Hartree-Fock theory*, Phys. Rev. Lett., 72 (1994), pp. 2981–2983.

[5] V. Bach, E. H. Lieb, and J. P. Solovej, *Generalized Hartree-Fock theory and the Hubbard model*, J. Statist. Phys., 76 (1994), pp. 3–89.

[6] S. Baroni and R. Resta, *Ab initio calculation of the macroscopic dielectric constant in silicon*, Phys. Rev. B, 33 (1986), pp. 7017–7021.

[7] C. Brouder, G. Panati, M. Calandra, C. Mourougane, and N. Marzari, *Exponential localization of Wannier functions in insulators*, Physical Review Letters, 98 (2007), p. 046402.

[8] É. Cancès, *Self-consistent field algorithms for Kohn–Sham models with fractional occupation numbers*, J. Chem. Phys., 114 (2001), p. 10616.

[9] É. Cancès, A. Deleurence, and M. Lewin, *A new approach to the modelling of local defects in crystals: the reduced Hartree-Fock case*, Commun. Math. Phys., 281 (2008), pp. 129–177.

[10] ———, *Non-perturbative embedding of local defects in crystalline materials*, J. Phys.: Condens. Matter, 20 (2008), p. 294213.

[11] É. Cancès and C. Le Bris, *Can we outperform the DIIS approach for electronic structure calculations?*, Int. J. Quantum Chem., 79 (2000), pp. 82–90.

[12] É. Cancès and M. Lewin, *The dielectric permittivity of crystals in the reduced Hartree-Fock approximation.* `arXiv:0903.1944`, 2009.

[13] I. Catto, C. Le Bris, and P.-L. Lions, *On the thermodynamic limit for Hartree-Fock type models*, Ann. Inst. H. Poincaré Anal. Non Linéaire, 18 (2001), pp. 687–760.

[14] R. Dovesi, R. Orlando, C. Roetti, C. Pisani, and V. Saunders, *The periodic Hartree-Fock method and its implementation in the Crystal code*, physica status solidi (b), 217 (2000), pp. 63–88.

[15] G. E. Engel and B. Farid, *Calculation of the dielectric properties of semiconductors*, Phys. Rev. B, 46 (1992), pp. 15812–15827.

[16] M. Gajdoš, K. Hummer, G. Kresse, J. Furthmüller, and F. Bechstedt, *Linear optical properties in the projector-augmented wave methodology*, Phys. Rev. B, 73 (2006), p. 045112.

[17] M. Ghimenti and M. Lewin, *Properties of periodic Hartree-Fock minimizers*, Calc. Var. Partial Differential Equations, 35 (2009), pp. 39–56.

[18] C. Hainzl, M. Lewin, and J. P. Solovej, *The mean-field approximation in quantum electrodynamics: the no-photon case*, Comm. Pure Appl. Math., 60 (2007), pp. 546–596.

[19] M. S. Hybertsen and S. G. Louie, *Ab initio static dielectric matrices from the density-functional approach. I. Formulation and application to semiconductors and insulators*, Phys. Rev. B, 35 (1987), pp. 5585–5601.

[20] ———, *Ab initio static dielectric matrices from the density-functional approach. II. Calculation of the screening response in diamond, Si, Ge, and LiCl*, Phys. Rev. B, 35 (1987), pp. 5602–5610.

[21] C. Kittel, *Quantum Theory of Solids*, John Wiley & Sons, 2nd ed., 1987.

[22] K. N. Kudin, G. E. Scuseria, and É. Cancès, *A black-box self-consistent field convergence algorithm: One step closer*, J. Chem. Phys., 116 (2002), p. 8255.

[23] K. Kunc and E. Tosatti, *Direct evaluation of the inverse dielectric matrix in semiconductors*, Phys. Rev. B, 29 (1984), pp. 7045–7047.

[24] E. H. Lieb, *Variational principle for many-fermion systems*, Phys. Rev. Lett., 46 (1981), pp. 457–459.

[25] E. H. Lieb, J. P. Solovej, and J. Yngvason, *Asymptotics of heavy atoms in high magnetic fields: I. Lowest Landau band regions*, Comm. Pure Appl. Math., 47 (1994), pp. 513–591.

[26] N. Marzari and D. Vanderbilt, *Maximally localized generalized wannier functions for composite energy bands*, Phys. Rev. B, 56 (1997), pp. 12847–12865.

[27] G. Nenciu, *Existence of the exponentially localised Wannier functions*, Commun. Math. Phys., 91 (1983), pp. 81–85.

[28] G. Panati, *Triviality of Bloch and Bloch-Dirac bundles*, Ann. Henri Poincaré, 8 (2007), pp. 995–1011.

[29] R. M. Pick, M. H. Cohen, and R. M. Martin, *Microscopic theory of force constants in the adiabatic approximation*, Phys. Rev. B, 1 (1970), pp. 910–920.

[30] C. Pisani, *Quantum-mechanical treatment of the energetics of local defects in crystals: A few answers and many open questions*, Phase Transitions, 52 (1994), pp. 123–136.

[31] M. Reed and B. Simon, *Methods of modern mathematical physics. IV. Analysis of operators*, Academic Press, New York, 1978.

[32] R. Resta and A. Baldereschi, *Dielectric matrices and local fields in polar semiconductors*, Phys. Rev. B, 23 (1981), pp. 6615–6624.

[33] C. Sgiarovello, M. Peressi, and R. Resta, *Electron localization in the insulating state: Application to crystalline semiconductors*, Phys. Rev. B, 64 (2001), p. 115202.

[34] A. M. Stoneham, *Theory of Defects in Solids: Electronic Structure of Defects in Insulators and Semiconductors*, Oxford University Press, 2001.

[35] L. E. Thomas, *Time dependent approach to scattering from impurities in a crystal*, Commun. Math. Phys., 33 (1973), pp. 335–343.

[36] G. H. Wannier, *The structure of electronic excitation levels in insulating crystals*, Phys. Rev., 52 (1937), pp. 191–197.

[37] N. Wiser, *Dielectric constant with local field effects included*, Phys. Rev., 129 (1963), pp. 62–69.

6

Spectral pollution and how to avoid it

This chapter is a summary of a work [18] with É. Séré. Let us consider a (bounded or unbounded, always densely defined) self-adjoint operator A acting on a separable Hilbert space $\mathfrak{H}$, with domain $D(A)$ and spectrum $\sigma(A)$. Let $\{V_n\}_{n\geq 1}$ be a sequence of finite-dimensional spaces such that $V_n \subset D(A)$, $V_n \subset V_{n+1}$, and $\overline{\cup_{n\geq 1} V_n} = \mathfrak{H}$. For each n, we may define $A_{|V_n} := P_{V_n} A P_{V_n}$ where P_{V_n} is the orthogonal projector on V_n. The operator $A_{|V_n}$ can just be seen as a $d_n \times d_n$ matrix where $d_n = \dim(V_n)$, hence the spectrum $\sigma(A_{|V_n})$ is just discrete. The goal of this chapter is to address the natural question whether one has $\sigma(A_{|V_n}) \to \sigma(A)$ in an appropriate sense.

To emphasize the possible difficulties, we give an example which already contains many of the general ideas. Let $\{e_n\}_{n\geq 1}$ be an orthonormal basis of $\mathfrak{H}$ and let

$$A := \sum_{n\geq 1} |e_{2n}\rangle\langle e_{2n}|, \qquad V_n = \operatorname{span}\big(e_1\,, \cdots, \,e_{2n}\,, \,\sin(\theta)e_{2n+1} + \cos(\theta)e_{2n+2}\big).$$

Note that A is simply the orthogonal projector on the subspace spanned by $\{e_{2n}\}$. Both its kernel and its range are infinite-dimensional, hence $\sigma(A) = \sigma_{\rm ess}(A) = \{0, 1\}$. The matrix of $A_{|V_n}$ is just

$$A_{|V_n} = \begin{pmatrix} 0 & & & & & \\ & 1 & & & & \\ & & 0 & & & \\ & & & 1 & & \\ & & & & \ddots & \\ & & & & & \cos^2(\theta) \end{pmatrix},$$

i.e. $\sigma(A_{|V_n}) = \{0, 1, \cos^2(\theta)\}$ for all n. We see that the spectrum of $A_{|V_n}$ contains, for all $n \geq 1$, an eigenvalue which has nothing to do with the true spectrum of A. Such an eigenvalue is called a *spurious eigenvalue* and the associated phenomenon is usually referred to as *spectral pollution* (see Figure 6.1).

Clearly, the above construction was possible because we have two eigenspaces of infinite multiplicity, i.e. because 0 and 1 are point of the *essential spectrum* of A. Also it is fairly easy to see that the above construction may be generalized in order to get any discrete set of polluted eigenvalues in $(0, 1)$, even a dense set in the limit $n \to \infty$. The purpose of this chapter is to study this phenomenon in details, from a rather abstract point of view and also with applications to periodic Schrödinger operators and to Dirac operators.

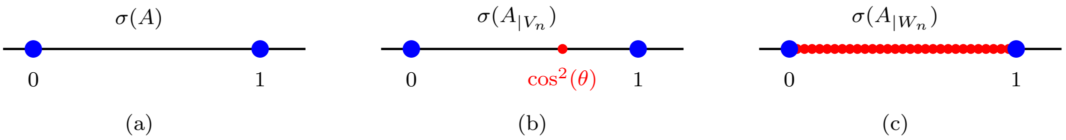


Figure 6.1: (a) True spectrum of the operator A. (b) Spectrum of our example $A_{|V_n}$, with one spurious eigenvalue. (c) It is possible to construct an example with a dense spectrum in $[0, 1]$.

6.1 A theorem of Levitin and Shargorodsky

In the whole chapter we consider a bounded or unbounded self-adjoint operator A on a separable Hilbert space $\mathfrak{H}$, with (dense) domain $D(A)$.

Definition 6.1 (Spurious eigenvalues)**.** *We say that* $\lambda \in \mathbb{R}$ *is a* spurious eigenvalue *of the operator* A *if there exists a sequence of finite dimensional spaces* $\{V_n\}_{n\geq 1}$ *with* $V_n \subset D(A)$ *and* $V_n \subset V_{n+1}$ *for any* n*, such that*

(i) $\overline{\cup_{n\geq 1} V_n}^{D(A)} = D(A)$*;*

(ii) $\lim_{n\to\infty} \operatorname{dist}\big(\lambda\,,\,\sigma(A_{|V_n})\big) = 0$*;*

(iii) $\lambda \notin \sigma(A)$*.*

We denote by $\operatorname{Spu}(A)$ *the set of spurious eigenvalues of* A*.*

In this chapter we shall only consider the spectral pollution issue, and we shall not study how well the spectrum $\sigma(A)$ of A is approximated by the discretized spectra $\sigma(A_{|V_n})$. Let us only mention that for every $\lambda \in \sigma(A)$, we have $\operatorname{dist}(\lambda, \sigma(A_{|V_n})) \to 0$ as $n \to \infty$, provided that (i) holds in Definition 6.1.

We will use like in [17] the notation $\hat{\sigma}_{\rm ess}(A)$ to denote the essential spectrum of A union $-\infty$ (and/or $+\infty$) if there exists a sequence of $\sigma(A) \ni \lambda_n \to -\infty$ (and/or $+\infty$). The following was proved by Levitin and Shargorodsky [17]:

Theorem 6.1 (Pollution in all spectral gaps [17])**.** *Let* A *be a self-adjoint operator on* $\mathfrak{H}$ *with dense domain* $D(A)$*. Then*

$$\overline{\operatorname{Spu}(A)} \cup \hat{\sigma}_{\rm ess}(A) = \operatorname{Conv}\left(\hat{\sigma}_{\rm ess}(A)\right).$$

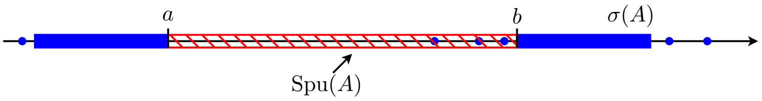


Figure 6.2: For an operator A which has a spectral gap $[a, b]$ in its essential spectrum, pollution can occur in the whole gap.

A similar result was previously obtained for bounded self-adjoint operators in [20] and generalized to bounded non self-adjoint operators in [7]. A short proof of Theorem 6.1 was given in [18]. All the main ideas are essentially contained the example we gave in the introduction. Levitin and Shargorodky have even shown that one can always construct one increasing sequence of finite-dimensional spaces $\{V_n\}$ which will create spurious eigenvalues in *all* gaps within the essential spectrum.

Indeed, spectral pollution is an important issue which arises in many different practical situations. It is encountered when approximating the spectrum of perturbations of periodic Schrödinger operators [4] or Strum-Liouville operators [25, 26, 1]. It is a very well reported difficulty in Quantum

Chemistry and Physics in particular regarding relativistic computations [10, 14, 16, 24, 11, 19, 23]. It also appears in elasticity, electromagnetism and hydrodynamics; see, e.g. the references in [2]. Eventually, it has raised as well a huge interest in the mathematical community, see, e.g., [17, 6, 4, 15, 7, 20, 21].

In [18] we have studied spectral pollution from a rather new perspective. Although many works focus on how to determine if an approximate eigenvalue is spurious or not (see, e.g., the rather successful second-order projection method [17, 4, 3]), we have on the contrary found conditions on the sequence $\{V_n\}$ which ensure that there will not be any pollution at all, in a given interval of the real line. This corresponds to some methods which are encountered in relativistic calculations [10, 14, 16, 24, 11, 19, 23]. We have indeed studied in details most of the techniques used by Chemists and Physicists in this setting, as we will explain.

6.2 Spectral pollution associated with a splitting of the ambient Hilbert space

The purpose of this section is to study spectral pollution if we add some assumptions on $\{V_n\}$. More precisely we will fix an orthogonal projector P acting on $\mathfrak{H}$ and we will add the natural assumption that P_{V_n} commute with P for all n, i.e. that V_n only contains vectors from $P\mathfrak{H}$ and $(1-P)\mathfrak{H}$. As we will see, under this new assumption the polluted spectrum (union $\hat{\sigma}_{\rm ess}(A)$) will in general be *the union of two intervals*. Saying differently, by adding such an assumption on $\{V_n\}$, we can *create a hole in the polluted spectrum*. Choosing P appropriately might then allow to avoid pollution in a given gap of the spectrum.

Note that our results of this section can easily be generalized to the case of a partition of unity $\{P_i\}_{i=1}^p$ of commuting projectors such that $1=\sum_{i=1}^p P_i$. Adding the assumption that P_{V_n} commutes with all P_i's, we would create p holes in the polluted spectrum. This might be useful if one wants to avoid spectral pollution in several gaps at the same time.

6.2.1 A general result

We start by defining properly P-spurious eigenvalues.

Definition 6.2 (Spurious eigenvalues associated with a splitting). *Consider an orthogonal projection $P:\mathfrak{H}\to\mathfrak{H}$. We say that $\lambda\in\mathbb{R}$ is a* P-spurious eigenvalue *of the operator A if there exist two sequences of finite dimensional spaces $\{V_n^+\}_{n\geq1}\subset P\mathfrak{H}\cap D(A)$ and $\{V_n^-\}_{n\geq1}\subset(1-P)\mathfrak{H}\cap D(A)$ with $V_n^\pm\subset V_{n+1}^\pm$ for any n, such that*

1. $\overline{\cup_{n\geq1}(V_n^-\oplus V_n^+)}^{D(A)}=D(A)$;
2. $\lim_{n\to\infty}\operatorname{dist}\left(\lambda,\sigma\left(A_{|(V_n^+\oplus V_n^-)}\right)\right)=0$;
3. $\lambda\notin\sigma(A)$.

We denote by $\operatorname{Spu}(A,P)$ *the set of P-spurious eigenvalues of the operator A.*

In [18] we have characterized exactly the spurious spectrum, in a similar way as in the theorem of Levitin and Shargorodsky:

Theorem 6.2 (Characterization of P-spurious eigenvalues [18]). *Let A be a self-adjoint operator with dense domain $D(A)$. Let P be an orthogonal projector on $\mathfrak{H}$ such that $PD(A)\subset D(A)$. We assume that PAP (resp. $(1-P)A(1-P)$) is essentially self-adjoint on $PD(A)$ (resp. $(1-P)D(A)$), with closure denoted as $A_{|P\mathfrak{H}}$ (resp. $A_{|(1-P)\mathfrak{H}}$). We assume also that*

$$\inf\hat{\sigma}_{\rm ess}\big(A_{|(1-P)\mathfrak{H}}\big)\leq\inf\hat{\sigma}_{\rm ess}\big(A_{|P\mathfrak{H}}\big).\tag{6.1}$$

Then we have

$$\overline{\mathrm{Spu}(A,P)} \cup \hat{\sigma}_{\mathrm{ess}}(A) = \left[\inf \hat{\sigma}_{\mathrm{ess}}(A), \sup \hat{\sigma}_{\mathrm{ess}}\left(A_{|(1-P)\mathfrak{H}}\right)\right] \cup \left[\inf \hat{\sigma}_{\mathrm{ess}}\left(A_{|P\mathfrak{H}}\right), \sup \hat{\sigma}_{\mathrm{ess}}(A)\right]. \tag{6.2}$$

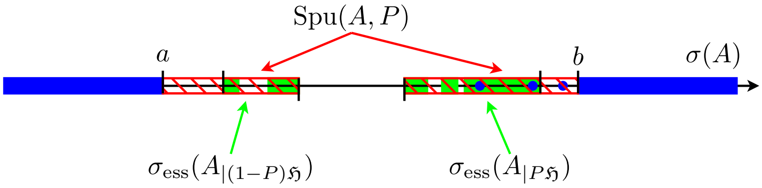


Figure 6.3: Illustration of Theorem 6.2: for an operator A with a gap $[a, b]$ in its essential spectrum, pollution can occur in the whole gap, except between the convex hulls of $\hat{\sigma}_{\mathrm{ess}}\left(A_{|P\mathfrak{H}}\right)$ and $\hat{\sigma}_{\mathrm{ess}}\left(A_{|(1-P)\mathfrak{H}}\right)$.

Let us emphasize that condition (6.1) always holds true, exchanging P and $1-P$ if necessary. Usually we will assume for convenience that $1-P$ is "associated with the lowest part of the spectrum" in the sense of (6.1).

The proof of Theorem 6.2 is more involved than the one of Theorem 6.1. Adapting ideas of Levitin and Shargorodsky, one easily sees that $\mathrm{Conv}(\hat{\sigma}_{\mathrm{ess}}(A_{|P\mathfrak{H}})) \cup \mathrm{Conv}(\hat{\sigma}_{\mathrm{ess}}(A_{|(1-P)\mathfrak{H}})) \subset \mathrm{Spu}(A, P)$ (i.e. the convex hull of the green sets on Figure 6.3). The difficult part is to get the other part of the polluted spectrum, and to show that there is no pollution between the two convex hulls.

An interesting example is when A possesses a gap $[a, b]$ in its essential spectrum, i.e. such that $(a, b) \cap \sigma_{\mathrm{ess}}(A) = \emptyset$, $[-\infty, a] \cap \hat{\sigma}_{\mathrm{ess}}(A) \neq \emptyset$ and $[b, \infty] \cap \hat{\sigma}_{\mathrm{ess}}(A) \neq \emptyset$. In this case it is easily seen that $\mathrm{Spu}(A, \Pi) \cap (a, b) = \emptyset$ for $\Pi := \chi_{[c,\infty)}(A)$ with $c = (a+b)/2$. The idea that we shall pursue in the next section is simply that if P is "not too far from Π", then we may be able to avoid completely pollution in the gap $[a, b]$.

6.2.2 A simple criterion of no pollution

Using Theorem 6.2, the following was proved in [18]:

Theorem 6.3 (Compact perturbations of spectral projector do not pollute [18])**.** *Let A be a self-adjoint operator defined on a dense domain $D(A)$, and let $a < b$ be such that*

$$(a, b) \cap \sigma_{\mathrm{ess}}(A) = \emptyset \quad \textit{and} \quad \mathrm{Tr}\left(\chi_{(-\infty,a]}(A)\right) = \mathrm{Tr}\left(\chi_{[b,\infty)}(A)\right) = +\infty. \tag{6.3}$$

Let $c \in (a, b) \setminus \sigma(A)$ and denote $\Pi := \chi_{(c,\infty)}(A)$. Let P be an orthogonal projector satisfying the assumptions of Theorem 6.2. We furthermore assume that $(P - \Pi)|A - c|^{1/2}$, initially defined on $D(|A - c|^{1/2})$, extends to a compact operator on $\mathfrak{H}$. Then we have

$$\mathrm{Spu}(A, P) \cap (a, b) = \emptyset.$$

Remark 6.1. *In [18], a simple example was given, showing that the power $1/2$ in $|A - c|^{1/2}$ is sharp: one can construct an (unbounded) operator A and a projection P such that $(P - \Pi)|A - c|^{\alpha}$ is compact for all $0 \leq \alpha < 1/2$ but $\mathrm{Spu}(A, P) \cap (a, b) \neq \emptyset$.*

The proof of Theorem 6.3 essentially consists in showing that one has, using the notation of Theorem 6.2, $\sigma_{\mathrm{ess}}(A_{|P\mathfrak{H}}) \subset [b, \infty)$ and $\sigma_{\mathrm{ess}}(A_{|(1-P)\mathfrak{H}}) \subset (-\infty; a]$.

The following simple application of the above result is useful:

Corollary 6.1. *Let A be a <u>bounded-below</u> self-adjoint operator defined on a dense domain $D(A)$, and let $a < b$ be such that*

$$(a, b) \cap \sigma_{\mathrm{ess}}(A) = \emptyset \quad \textit{and} \quad \mathrm{Tr}\left(\chi_{(-\infty,a]}(A)\right) = \mathrm{Tr}\left(\chi_{[b,\infty)}(A)\right) = +\infty. \tag{6.4}$$

Let $c \in (a, b)$ be such that $c \notin \sigma(A)$ and denote $\Pi := \chi_{(c,\infty)}(A)$.

Let B be a symmetric operator such that $A + B$ is self-adjoint on $D(A)$ and such that $\big((A + B - i)^{-1} - (A - i)^{-1}\big)|A - c|^{1/2}$, initially defined on $D(|A - c|^{1/2})$, extends to a compact operator on $\mathfrak{H}$. Then we have

$$\mathrm{Spu}(A + B, \Pi) \cap (a, b) = \emptyset.$$

Remark 6.2. *Again the power $1/2$ in $|A - c|^{1/2}$ is optimal, as shown in [18]. Corollary 6.1 is* a priori *wrong when A is not semi-bounded. One can construct operators A and B such that (with A not semi-bounded) such that $\big((A + B + i)^{-1} - (A + i)^{-1}\big)|A|^{\alpha}$ is compact for all $0 \le \alpha < 1$ whereas* $\mathrm{Spu}(A + B, \Pi) \cap (a, b) \neq \emptyset$. *Hence in the case of not semi-bounded operators, a reasonable conjecture is that Corollary 6.1 holds true with $1/2$ replaced by 1, but we have been unable to prove this.*

6.2.3 Applications

6.2.3.1 Periodic Schrödinger operators

In this section, we show that approximating eigenvalues in gaps of periodic Schrödinger operators using a so-called *Wannier basis* as was explained in Chapter 5, Section 5.3 (following [5]), does not yield any spurious eigenvalue. For references on spectral pollution in this setting, we refer for example to [4].

Consider a potential V_{per} which is $\mathbb{Z}^d$-periodic in $\mathbb{R}^d$ (the generalization to any lattice is straightforward and tackled in [18]) such that

$$V_{\mathrm{per}} \in L^p_{\mathrm{loc}}(\mathbb{R}^d) \quad \text{where} \quad \begin{cases} p = 2 & \text{if } d \le 3, \\ p > 2 & \text{if } d = 4, \\ p = d/2 & \text{if } d \ge 5, \end{cases}$$

and a decaying potential

$$W \in L^q(\mathbb{R}^d) \cap L^p_{\mathrm{loc}}(\mathbb{R}^d) + L^\infty_\epsilon(\mathbb{R}^d)$$

for some $q > \max(d/3, 1)$ and p as before. As is well-known [22], the operators

$$A_{\mathrm{per}} := -\Delta + V_{\mathrm{per}}, \qquad A = A_{\mathrm{per}} + W$$

are self-adjoint on $H^2(\mathbb{R}^d)$. The spectrum of A_{per} is composed of bands as seen via the Bloch-Floquet decomposition recalled in Chapter 5. When V_{per} is smooth enough [29, 22], the spectrum of A_{per} is purely absolutely continuous. Under our assumptions on W, $(A_{\mathrm{per}} + W - i)^{-1} - (A_{\mathrm{per}} - i)^{-1}$ is $(1 - \Delta)^{-1/2}$-compact as seen by the resolvent expansion [22], and one has $\sigma_{\mathrm{ess}}(A) = \sigma(A_{\mathrm{per}})$.

Using the Bloch-Floquet decomposition, a spectral decomposition of the reference periodic operator A_{per} is easily accessible numerically. This decomposition can be used as a starting point to avoid pollution for the perturbed operator A. For simplicity we shall assume that the spectral decomposition of A_{per} is known exactly. More precisely we make the assumption that there is a gap (a, b) between the kth and the $(k + 1)$st band and that the associated spectral projector $P_{\mathrm{per}} := \chi_{(-\infty,c)}(A_{\mathrm{per}})$ with $c = (a + b)/2$ is known. The interest of this approach is the following immediate consequence of Corollary 6.1:

Theorem 6.4 (No pollution for periodic Schrödinger operators). *We assume V_{per} and W are as before. Then we have*

$$\mathrm{Spu}(A, P_{\mathrm{per}}) \cap (a, b) = \emptyset. \tag{6.5}$$

As was noticed in [5] and re-explained in Chapter 5, Section 5.3, a natural basis respecting the spectral decomposition of A_{per} is given by a so-called *Wannier basis*.

6.2.3.2 Dirac operator in upper/lower spinor basis

An important problem is to compute eigenvalues of operators of the form

$$D^V = D^0 + V$$

where D^0 is the usual Dirac operator introduced in Chapter 4,

$$D^0 = -ic\sum_{k=1}^{3}\alpha_k\partial_{x_k} + \beta mc^2$$

(α_k and β are the Pauli matrices defined in (4.5)) and V is a multiplication operator by a real function $x \mapsto V(x)$. In practice, spectral pollution is an important problem [10, 14, 16, 24] which is dealt with in Quantum Physics and Chemistry by means of several different methods, the most widely used being the so-called *kinetic balance* which we will study later in Section 6.3. We refer to [3] for a recent numerical study based on the so-called *second-order method* for the radial Dirac operator.

We now present a heuristic argument which can be made mathematically rigorous in many cases [28, 12]. First we write the equation satisfied by an eigenvector (φ, χ) of $D^0 + V$ with eigenvalue $mc^2 + \lambda \in (-mc^2, mc^2)$ as follows:

$$\begin{cases} (mc^2 + V)\varphi + c\sigma\cdot(-i\nabla)\chi = (mc^2+\lambda)\varphi, \\ (-mc^2 + V)\chi + c\sigma\cdot(-i\nabla)\varphi = (mc^2+\lambda)\chi, \end{cases} \tag{6.6}$$

where we recall that $\sigma = (\sigma_1, \sigma_2, \sigma_3)$ are the Pauli matrices defined in Chapter 4, Equation (4.6). Hence one deduces that (when it makes sense)

$$\chi = \frac{c}{2mc^2 + \lambda - V}\sigma\cdot(-i\nabla)\varphi. \tag{6.7}$$

If V and λ stay bounded, we infer that, at least formally,

$$\begin{pmatrix} \varphi \\ \chi \end{pmatrix} \sim_{c\to\infty} \begin{pmatrix} \varphi \\ \frac{1}{2mc}\sigma\cdot(-i\nabla)\varphi \end{pmatrix}. \tag{6.8}$$

Hence we see that in the nonrelativistic limit $c \to \infty$, the eigenvectors of A associated with a positive eigenvalue converge to a vector of the form $\begin{pmatrix}\varphi\\0\end{pmatrix}$. Reintroducing the asymptotic formula (6.8) of χ in the first equation of (6.6), one gets that φ is an eigenvector of the nonrelativistic operator $-\Delta/(2m) + V$ in $L^2(\mathbb{R}^3, \mathbb{C}^2)$.

For this reason, it is very natural to consider a splitting of the Hilbert space $L^2(\mathbb{R}^3, \mathbb{C}^4)$ into *upper* and *lower spinor* and we introduce the following orthogonal projector

$$\boxed{\mathcal{P}\begin{pmatrix}\varphi\\\chi\end{pmatrix} = \begin{pmatrix}\varphi\\0\end{pmatrix}, \qquad \varphi, \chi \in L^2(\mathbb{R}^3, \mathbb{C}^2).} \tag{6.9}$$

This splitting is the choice of most of the methods we are aware of in Quantum Physics and Chemistry. Applying Theorem 6.2, we can characterize the spurious spectrum associated with this splitting. For simplicity we take $m = c = 1$ in the following.

Theorem 6.5 (Pollution in upper/lower spinor basis for Dirac operators [18])**.** *Assume that the real function V satisfies the following assumptions:*

(i) there exist $\{R_k\}_{k=1}^M \subset \mathbb{R}^3$ and a positive number $r < \inf_{k\neq\ell}|R_k - R_\ell|/2$ such that

$$\max_{k=1..K}\sup_{|x-R_k|\le r}|x - R_k|\,|V(x)| < \frac{\sqrt{3}}{2}; \tag{6.10}$$

(ii) one has

$$V\mathbb{1}_{\mathbb{R}^3\setminus\cup_1^K B(R_k,r)} \in L^p(\mathbb{R}^3) + L^\infty_\epsilon(\mathbb{R}^3) \qquad \text{for some } 3 < p < \infty. \tag{6.11}$$

Let $\mathcal{P}$ *be as in* (6.9)*. Then one has*

$$\overline{\mathrm{Spu}(D^0+V,\mathcal{P})} = \Big\{\mathrm{Conv}\,\big(\mathrm{Ess}\big(1+V\big)\big)\cup\mathrm{Conv}\,\big(\mathrm{Ess}\big(-1+V\big)\big)\Big\}\cap[-1,1] \tag{6.12}$$

where $\mathrm{Ess}(W)$ *denotes the essential range of the function* W*, i.e.*

$$\mathrm{Ess}(W) = \Big\{\lambda\in\mathbb{R}\ \big|\ \big|W^{-1}([\lambda-\epsilon,\lambda+\epsilon])\big| \neq 0\ \forall\epsilon>0\Big\}.$$

Our assumptions on V cover the case of the Coulomb potential, $V(x) = \kappa|x|^{-1}$ when $|\kappa| < \sqrt{3}/2$. In our units, this corresponds to nuclei which have less than 118 protons, which covers all existing atoms. On the other hand, a typical example for which $V \in L^p(\mathbb{R}^3)\cap L^\infty(\mathbb{R}^3)$ is the case of smeared nuclei $V = \rho * 1/|x|$ where ρ is a (sufficiently smooth) distribution of charge for the nuclei.

As it may be seen, using the characterization of eigenvalues in gaps which was provided by Dolbeault, Esteban and Séré in [9], the discrete spectrum of D^0+V lies precisely in the set $\Big\{\mathrm{Conv}\,\big(\mathrm{Ess}\big(1+V\big)\big)\cup\mathrm{Conv}\,\big(\mathrm{Ess}\big(-1+V\big)\big)\Big\}\cap[-1,1]$. Therefore Theorem 6.5 shows that pollution can in principle occur in all interesting places. Physicists and Chemists then add constraints on the basis to avoid this phenomenon, like the kinetic balance method which will be studied later in Section 6.3.

6.2.3.3 Dirac operator in free basis

In this section, we prove that a way to avoid pollution in the *whole gap* is to take a basis associated with the spectral decomposition of the free Dirac operator, i.e. choosing as projector $P^0_+ := \chi_{(0,\infty)}(D^0)$. As we will see this choice does not rely on the size of V like in the previous section. Its main disadvantage is that constructing a basis preserving the decomposition induced by P^0_+ requires a Fourier transform, which might increase the computational cost dramatically. However this might be the only reasonable option to describe strong QED-like effects. In [18], another method proposed by Shabaev et al was also studied but we will not detail this here.

Theorem 6.6 (No pollution in free basis [18])**.** *Assume that* V *is a real function such that*[1]

$$V\in L^p(\mathbb{R}^3) + \Big(L^r(\mathbb{R}^3)\cap\dot{W}^{1,q}(\mathbb{R}^3)\Big) + L^\infty_\epsilon(\mathbb{R}^3) \tag{6.13}$$

for some $6 < p < \infty$*, some* $3 < r \leq 6$ *and some* $2 < q < \infty$*, or that*

$$V = \kappa|x|^{-1}$$

for some $|\kappa| < \sqrt{3}/2$*. Then one has*

$$\mathrm{Spu}(D^0+V,P^0_+) = \emptyset.$$

The awkward condition (6.13) was chosen in such a way that the operator $\big((D^0+V+i)^{-1} - (D^0+i)^{-1}\big)|D^0|^{1/2}$ is compact, as seen by the resolvent expansion. The result in this case is then a consequence of Theorem 6.3. In the Coulomb case the corresponding operator is not compact but the essential spectra of $P^0_+(D^0+V)P^0_+$ and $P^0_-(D^0+V)P^0_-$ are known from [13] and it then suffices to apply directly Theorem 6.2.

---

[1] We have used the notation $L^r(\mathbb{R}^3)\cap\dot{W}^{1,q}(\mathbb{R}^3) = \{V\in L^r(\mathbb{R}^3)\ |\ \nabla V\in L^q(\mathbb{R}^3)\}$.

6.3 Balanced basis

We have seen in Theorem 6.5 that a basis made of upper and lower spinors can give rise to pollution. In Chemistry and Physics, the usual solution is the so-called *kinetic balanced method*, which consists in imposing a specific relation between the vectors in the upper and lower spinor basis sets. The main argument consists in noticing from (6.8) that the lower spinor basis set must be able to reproduce correctly the derivative of the functions used to form the upper spinor basis. This leads to choosing the lower spinor basis as $\{(\sigma\cdot p)\,\varphi_n\}$ when $\{\varphi_n\}$ is the chosen basis for the upper spinor.

As before we start by defining abstractly the concept of balanced basis before turning to applications in the Dirac case. For simplicity we will not state all the abstract results of [18].

6.3.1 Abstract setting

Consider an orthogonal projection $P:\mathfrak{H}\to\mathfrak{H}$. Let $L:D(L)\subset P\mathfrak{H}\to(1-P)\mathfrak{H}$ be a (possibly unbounded) operator which we call *balanced operator.* We assume that

- L is an injection: if $Lx=0$ for $x\in D(L)$, then $x=0$;
- $D(L)\oplus LD(L)$ is a core for A.

Definition 6.3 (Spurious eigenvalues in balanced basis)**.** *We say that $\lambda\in\mathbb{R}$ is a (P,L)*-spurious eigenvalue *of the operator A if there exists a sequence of finite dimensional spaces $\{V_n^+\}_{n\geq1}\subset D(L)$ with $V_n^+\subset V_{n+1}^+$ for all n, such that*

1. $\overline{\cup_{n\geq1}(V_n^+\oplus LV_n^+)}^{D(A)}=D(A)$;
2. $\lim_{n\to\infty}\operatorname{dist}\left(\lambda,\sigma\left(A_{|(V_n^+\oplus LV_n^+)}\right)\right)=0$;
3. $\lambda\notin\sigma(A)$.

We denote by $\mathrm{Spu}(A,P,L)$ *the set of (P,L)-spurious eigenvalues of the operator A.*

For any fixed $0\neq x\in D(L)$, we consider the 2×2 matrix $M(x)$ of A restricted to the 2-dimensional space $x\mathbb{C}\oplus Lx\,\mathbb{C}$, and we denote by $\mu_1(x)\leq\mu_2(x)$ its eigenvalues. The vague idea that one should have is that the polluted spectrum will arise from accumulation points of the two functions $\mu_i(x_n^+)$ for sequences $\{x_n^+\}\subset D(L)$ converging weakly to zero.

Clearly the introduction of the operator L breaks the symmetry between the two spaces $P\mathfrak{H}$ and $(1-P)\mathfrak{H}$. For this reason we shall concentrate on pollution occurring in the *upper part of the spectrum* and we will not give necessary conditions for the lower part[2]. Let us introduce

$$d:=\sup\sigma(A_{(1-P)\mathfrak{H}}). \tag{6.14}$$

and assume that $d<\infty$. In [18] we have only studied (P,L)-spurious eigenvalues in (d,∞). Note that due to Theorem 6.2, it would be more natural to let instead $d:=\sup\hat{\sigma}_{\rm ess}(A_{(1-P)\mathfrak{H}})$ but this will actually not change anything for the examples we are interested in: in the Dirac case D^0+V and for $P=\mathcal{P}$, the orthogonal projector on the upper spinor defined in (6.9), the spectrum of $(D^0+V)_{|(1-P)L^2(\mathbb{R}^3,\mathbb{C}^4)}=-1+V$ is only composed of essential spectrum.

In [18], we could not characterize completely the (P,L)-spurious spectrum of A and we have provided some necessary or sufficient conditions. We will not explain this in details here and we only quote the following result, contained in [18]:

---

[2] As we have mentionned before we always assume for simplicity that $\inf\hat{\sigma}_{\rm ess}(A_{|(1-P)\mathfrak{H}})\leq\inf\hat{\sigma}_{\rm ess}(A_{|P\mathfrak{H}})$, i.e. that $1-P$ is responsible from the pollution occurring in the lower part of the spectrum.

Theorem 6.7 ((P,L)-spurious eigenvalues, necessary condition [18])**.** *Let be A, P and L as before. Let us define*

$$m = \inf_{\substack{\{x_n^+\}\subset D(L)\setminus\{0\},\\ x_n^+\rightharpoonup 0,\ \|x_n^+\|=1}} \liminf_{n\to\infty} \mu_2(x_n^+). \tag{6.15}$$

and assume that $m > d$ where d was defined in (6.14)*. We also assume that the following additional* continuity property *holds for some $b > d$:*

$$\left.\begin{array}{r} \{x_n^+\}\subset D(L) \\ x_n^+ \rightharpoonup 0 \\ \limsup\limits_{n\to\infty} \mu_2(x_n^+) < b \end{array}\right\} \Longrightarrow \langle Ax_n^+, x_n^+\rangle \to 0. \tag{6.16}$$

Then we have

$$\mathrm{Spu}(A,P,L)\cap \big(d, \min(m,b)\big) = \emptyset.$$

If $A_{|P\mathfrak{H}}$ is bounded, then condition (6.16) holds true for $b = +\infty > d$.

Theorem 6.7 has many similarities with the characterization of eigenvalues in a gap which was proved by Dolbeault, Esteban and Séré in [9] (where our number $d = \sup\sigma(A_{(1-P)\mathfrak{H}})$ was denoted by 'a'). In particular the reader should compare the assumptions $d < m$ with (iii) at the bottom of p. 209 in [9]. The proof of [18] indeed uses many ideas of [9], which was itself inspired by an important Physics paper of Talman [27] (he introduced a minimax principle for the Dirac equation in order to avoid spectral pollution).

6.3.2 Kinetic and Atomic Balance in Dirac calculations

We come back to the Dirac operator $A = D^0 + V$ for a potential satisfying the assumptions (6.10) and (6.11) of Theorem 6.5 and

$$\sup(V) < 2 \tag{6.17}$$

in such a way that $d = -1 + \sup(V) = \sup\sigma_{\rm ess}(A_{|(1-\mathcal{P})L^2})$ does not reach the top of the gap. We will indeed for simplicity concentrate ourselves on the case for which either V is bounded, or V is a purely attractive Coulomb potential, $V(x) = -\kappa/|x|$, $0 < \kappa < \sqrt{3}/2$. We take $P = \mathcal{P}$, the projector on the upper spinors.

The most common method is the so-called *kinetic balance* [10, 14, 16, 24]. It consists in choosing as balanced operator

$$\boxed{L_{KB} = -i\sigma\cdot\nabla} \tag{6.18}$$

We can for instance define L_{KB} on the domain $D(L_{KB}) = C_0^\infty(\mathbb{R}^3,\mathbb{C}^2)$, in which case L_{KB} satisfies all the previous assumptions. In [18], it was proved that kinetic balance allows to *avoid pollution in the upper part of the gap for smooth potentials*, hence for instance for $V = -\rho * |x|^{-1}$ where $\rho \geq 0$ is the distribution of charge for smeared nuclei. However, the kinetic balance method *does not avoid spectral pollution in the case of pointwise nuclei* (Coulomb potential):

Theorem 6.8 (Kinetic Balance [18])**.** ***(i) Bounded potential.*** *Assume that* $V \in L^p(\mathbb{R}^3)\cap L^\infty(\mathbb{R}^3)$ *for some* $p > 3$*, that* $\lim_{|x|\to\infty} V(x) = 0$*, and that* $\sup(V) < 2$ *Then we have*

$$\overline{\mathrm{Spu}(D^0 + V, \mathcal{P}, L_{KB})} = [-1, -1 + \sup V].$$

(ii) Coulomb potential. *Assume that* $0 < \kappa < \sqrt{3}/2$*. Then we have*

$$\overline{\mathrm{Spu}\left(D^0 - \frac{\kappa}{|x|}, \mathcal{P}, L_{KB}\right)} = [-1, 1]. \tag{6.19}$$

It seems a well-known fact in Quantum Chemistry and Physics [11, 19] that the kinetic balance method consisting in choosing $L = L_{KB}$ is not well-behaved for pointwise nuclei. The reason is that the behavior at zero of $c(2mc^2+\lambda-V)^{-1}\sigma\cdot(-i\nabla)$ appearing in (6.7), is not properly captured by $\sigma\cdot(-i\nabla)$, if $V(x) = -\kappa|x|^{-1}$. To better capture the behavior at zero, we study another method which we have called *atomic balance*. It consists in taking

$$\boxed{L_{AB} = \frac{1}{2-V}\sigma\cdot(-i\nabla)} \tag{6.20}$$

where we recall that we have assumed $2 > \sup(V)$. Provided that V is smooth enough, we can define L_{AB} on the domain $D(L_{AB}) = C_0^\infty(\mathbb{R}^3\setminus\{0\},\mathbb{C}^2)$, in which case L_{AB} satisfies all the previous assumptions. The following was proved in [18]:

Theorem 6.9 (Atomic Balance [18])**.** *Let V be such that* $\sup(V) < 2$, $(2-V)^{-2}\nabla V \in L^\infty(\mathbb{R}^3)$ *and*

$$-\frac{\kappa}{|x|} \le V(x)$$

for some $0 \le \kappa < \sqrt{3}/2$. *We also assume that the positive part* $\max(V,0)$ *is in* $L^p(\mathbb{R}^3)$ *for some* $p > 3$ *and that* $\lim_{|x|\to\infty} V(x) = 0$. *Then we have*

$$\overline{\mathrm{Spu}(D^0+V,\mathcal{P},L_{AB})} = [-1,-1+\sup V].$$

We now quickly explain the vague idea of the proof of Theorems 6.8 and 6.9. The number $\mu_2(\varphi)$ (the highest eigenvalue of D^0+V in the basis $\{\varphi, L\varphi\}$) is the largest solution to the following equation [9]

$$\langle(1+V)\varphi,\varphi\rangle + \frac{\big(\Re\langle L\varphi,\sigma\cdot(-i\nabla)\varphi\rangle\big)^2}{\langle(\mu+1-V)L\varphi,L\varphi\rangle} = \mu\,\|\varphi\|^2 \tag{6.21}$$

where the denominator of the second term does not vanish when $\mu_2(\varphi) > d = \sup(V)-1$. Note the term on the left is decreasing with respect to μ, whereas the term on the right is increasing with respect to μ. The idea is, using the above characterization, to show that $m \ge 1$, where m is defined in Theorem 6.7, see (6.15) (we will not explain why (6.16) holds in our cases). Also, when $m < 1$, one can usually construct by hand spurious eigenvalues in $[m,1]$. In the Coulomb case for the kinetic balance method, this is done by taking a sequence $\{\varphi_n\}$ composed of two bubbles concentrating at zero.

In the kinetic balance case of Theorem 6.8 (when V is bounded), arguing by contradiction leads to the existence of $\{\varphi_n\}$ with $\|\varphi_n\| = 1$ and $\varphi_n \to 0$, such that $\mu_2(\varphi_n) \to \mu \in (d,1)$. Using (6.21) this means

$$\int_{\mathbb{R}^3} V|\varphi_n|^2 + \frac{\left(\displaystyle\int_{\mathbb{R}^3}|\sigma\cdot\nabla\varphi_n|^2\right)^2}{\displaystyle\int_{\mathbb{R}^3}(\mu_2(\varphi_n)+1-V)|\sigma\cdot\nabla\varphi_n|^2} = \mu_2(\varphi_n) - 1 \to \mu - 1 < 0. \tag{6.22}$$

One can then see that $\{\varphi_n\}$ is bounded in H^1 and easily get a contradiction.

The atomic balance case is more involved. This time one gets a sequence satisfying

$$\int_{\mathbb{R}^3} V|\varphi_n|^2 + \frac{\left(\displaystyle\int_{\mathbb{R}^3}\frac{|\sigma\cdot\nabla\varphi_n|^2}{2-V}\right)^2}{\displaystyle\int_{\mathbb{R}^3}\frac{(\mu_2(\varphi_n)+1-V)|\sigma\cdot\nabla\varphi_n|^2}{(2-V)^2}} = \mu_2(\varphi_n) - 1 \to \mu - 1 < 0. \tag{6.23}$$

Using that $1+\mu-V \le 2-V$, one gets for n large enough

$$-\int_{\mathbb{R}^3}\frac{\kappa}{|x|}|\varphi_n|^2 + \left(\int_{\mathbb{R}^3}\frac{|\sigma\cdot\nabla\varphi_n|^2}{2+\kappa/|x|}\right)^2 \le \int_{\mathbb{R}^3} V|\varphi_n|^2 + \left(\int_{\mathbb{R}^3}\frac{|\sigma\cdot\nabla\varphi_n|^2}{2-V}\right)^2 \le \frac{\mu-1}{2} < 0. \tag{6.24}$$

One then gets a contradiction by splitting φ_n in three pieces, isolating the part at zero and at infinity, and using the following Hardy-type inequality:

$$\int_{\mathbb{R}^3} \frac{c^2|\sigma\cdot\nabla\varphi(x)|^2}{c^2+\frac{\nu}{|x|}+\sqrt{c^4-\nu^2c^2}}dx + (c^2-\sqrt{c^4-\nu^2c^2})\int_{\mathbb{R}^3}|\varphi(x)|^2dx \geq \nu\int_{\mathbb{R}^3}\frac{|\varphi(x)|^2}{|x|}dx. \tag{6.25}$$

This inequality was obtained in [9] by using a min-max characterization of the first eigenvalue of $-ic\boldsymbol{\alpha}\cdot\nabla + c^2\beta - \nu/|x|$. Indeed (6.25) is an equality when φ is equal to the upper spinor of the eigenfunction corresponding to the first eigenvalue in $(-c^2, c^2)$ of $-ic\boldsymbol{\alpha}\cdot\nabla + c^2\beta - \nu/|x|$. The inequality (6.25) was then proved by a direct analytical method in [8]. We note that, introducing $m = c(1+\sqrt{1-(\nu/c)^2})$ and $\kappa = \nu/c$ we can rewrite (6.25) in the following form

$$\int_{\mathbb{R}^3}\frac{|\sigma\cdot\nabla\varphi(x)|^2}{m+\frac{\kappa}{|x|}}dx + m\frac{1-\sqrt{1-\kappa^2}}{1+\sqrt{1-\kappa^2}}\int_{\mathbb{R}^3}|\varphi(x)|^2dx \geq \kappa\int_{\mathbb{R}^3}\frac{|\varphi(x)|^2}{|x|}dx. \tag{6.26}$$

References

[1] L. Aceto, P. Ghelardoni, and M. Marletta, *Numerical computation of eigenvalues in spectral gaps of Sturm-Liouville operators*, J. Comput. Appl. Math., 189 (2006), pp. 453–470.

[2] L. Boulton, *Non-variational approximation of discrete eigenvalues of self-adjoint operators*, IMA J. Numer. Anal., 27 (2007), pp. 102–121.

[3] L. Boulton and N. Boussaid, *Non-variational computation of the eigenstates of Dirac operators with radially symmetric potentials.* arXiv:0808.0228, 2008.

[4] L. Boulton and M. Levitin, *On approximation of the eigenvalues of perturbed periodic Schrödinger operators*, J. Phys. A, 40 (2007), pp. 9319–9329.

[5] É. Cancès, A. Deleurence, and M. Lewin, *Non-perturbative embedding of local defects in crystalline materials*, J. Phys.: Condens. Matter, 20 (2008), p. 294213.

[6] E. Davies and M. Plum, *Spectral pollution*, IMA J. Numer. Anal., 24 (2004), pp. 417–438.

[7] J. Descloux, *Essential numerical range of an operator with respect to a coercive form and the approximation of its spectrum by the galerkin method*, SIAM J. Numer. Anal., 18 (1981), pp. 1128–1133.

[8] J. Dolbeault, M. J. Esteban, M. Loss, and L. Vega, *An analytical proof of Hardy-like inequalities related to the Dirac operator*, J. Funct. Anal., 216 (2004), pp. 1–21.

[9] J. Dolbeault, M. J. Esteban, and É. Séré, *On the eigenvalues of operators with gaps. Application to Dirac operators*, J. Funct. Anal., 174 (2000), pp. 208–226.

[10] G. W. F. Drake and S. P. Goldman, *Application of discrete-basis-set methods to the dirac equation*, Phys. Rev. A, 23 (1981), pp. 2093–2098.

[11] K. G. Dyall and K. Fægri Jr, *Kinetic balance and variational bounds failure in the solution of the Dirac equation in a finite Gaussian basis set*, Chem. Phys. Letters, 174 (1990), pp. 25–32.

[12] M. J. Esteban, M. Lewin, and É. Séré, *Variational methods in relativistic quantum mechanics*, Bull. Amer. Math. Soc. (N.S.), 45 (2008), pp. 535–593.

[13] W. D. Evans, P. Perry, and H. Siedentop, *The spectrum of relativistic one-electron atoms according to Bethe and Salpeter*, Commun. Math. Phys., 178 (1996), pp. 733–746.

[14] I. P. Grant, *Conditions for convergence of variational solutions of Dirac's equation in a finite basis*, Phys. Rev. A, 25 (1982), pp. 1230–1232.

[15] A. C. Hansen, *On the approximation of spectra of linear operators on Hilbert spaces*, J. Funct. Anal., 254 (2008), pp. 2092–2126.

[16] W. Kutzelnigg, *Basis set expansion of the dirac operator without variational collapse*, Int. J. Quantum Chemistry, 25 (1984), pp. 107–129.

[17] M. Levitin and E. Shargorodsky, *Spectral pollution and second-order relative spectra for self-adjoint operators*, IMA J. Numer. Anal., 24 (2004), pp. 393–416.

[18] M. Lewin and É. Séré, *Spectral pollution and how to avoid it (with applications to Dirac and periodic Schrödinger operators).* `arXiv:0812.2153`, 2008.

[19] G. Pestka, *Spurious roots in the algebraic Dirac equation*, Phys. Scr., 68 (2003), pp. 254–258.

[20] A. Pokrzywa, *Method of orthogonal projections and approximation of the spectrum of a bounded operator*, Studia Math., 65 (1979), pp. 21–29.

[21] ———, *Method of orthogonal projections and approximation of the spectrum of a bounded operator. II*, Studia Math., 70 (1981), pp. 1–9.

[22] M. Reed and B. Simon, *Methods of modern mathematical physics. IV. Analysis of operators*, Academic Press, New York, 1978.

[23] V. Shabaev, I. I. Tupitsyn, V. A. Yerokhin, G. Plunien, and G. Soff, *Dual kinetic balance approach to basis-set expansions for the Dirac equation*, Phys. Rev. Lett., 93 (2004), p. 130405.

[24] R. E. Stanton and S. Havriliak, *Kinetic balance: A partial solution to the problem of variational safety in Dirac calculations*, J. Chem. Phys., 81 (1984), pp. 1910–1918.

[25] G. Stolz and J. Weidmann, *Approximation of isolated eigenvalues of ordinary differential operators*, J. Reine Angew. Math., 445 (1993), pp. 31–44.

[26] ———, *Approximation of isolated eigenvalues of general singular ordinary differential operators*, Results Math., 28 (1995), pp. 345–358.

[27] J. D. Talman, *Minimax principle for the Dirac equation*, Phys. Rev. Lett., 57 (1986), pp. 1091–1094.

[28] B. Thaller, *The Dirac equation*, Texts and Monographs in Physics, Springer-Verlag, Berlin, 1992.

[29] L. E. Thomas, *Time dependent approach to scattering from impurities in a crystal*, Commun. Math. Phys., 33 (1973), pp. 335–343.

A

Density matrices of Bogoliubov states having finite charge

This chapter provides some important tools of functional analysis, which we have used throughout the text. We work in an abstract separable Hilbert space $\mathfrak{H}$ and consider a fixed projector Π such that both Π and $1-\Pi$ have an infinite rank. The physical picture is that Π is a reference Hartree-Fock state representing our system in some unperturbed situation. The infinite rank of Π means that our system contains infinitely many particles.

In Chapter 2, we have

$$\Pi = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \mathfrak{H} = L^2(\mathbb{R}^3, \mathbb{C}^q) \oplus \mathcal{J} L^2(\mathbb{R}^3, \mathbb{C}^q)$$

where $\mathcal{J}$ is the conjugation operator.

In Chapter 4, we have $\mathfrak{H} = L^2(\mathbb{R}^3, \mathbb{C}^4)$ and $\Pi = \mathcal{P}^0_- = \chi_{(-\infty,0)}(\mathcal{D}^0)$ (Bogoliubov-Dirac-Fock model) or $\Pi = P^0_- = \chi_{(-\infty,0)}(D^0)$ (reduced Bogoliubov-Dirac-Fock model) which represent the free vacuum (Dirac sea) in the absence of any external source, in the corresponding model.

In Chapter 5, we have $\mathfrak{H} = L^2(\mathbb{R}^3, \mathbb{C})$ and $\Pi = P^0_{\rm per} = \chi_{(-\infty,\epsilon_{\rm F})}(H^0_{\rm per})$ which is the (reduced) Hartree-Fock state of the Fermi sea of the perfect crystal, in the absence of any defect.

When describing perturbations of a quantum system around the state Π, we had to consider the following set

$$\mathcal{K} := \left\{ Q \in \mathfrak{S}_1^\Pi(\mathfrak{H}) \mid Q^* = Q, \ -\Pi \leq Q \leq 1 - \Pi \right\}. \tag{A.1}$$

where the definition of $\mathfrak{S}_1^\Pi(\mathfrak{H})$ will be recalled below. The set $\mathcal{K}$ can indeed be defined as the (closure of the) convex hull of operators of the form $Q = P - \Pi$ where P is an orthogonal projector and $P - \Pi$ is Hilbert-Schmidt. Any such projection P represents another Hartree-Fock state which is not too far from the reference state Π. As we will see below, the interpretation is that any $Q \in \mathcal{K}$ is uniquely associated with a quasi-free state having a finite charge, in the Fock representation associated with Π. States corresponding to $Q = P - \Pi$ with P being a projector are usually called *Bogoliubov states* because they can be obtained by applying a Bogoliubov rotation to the reference state Π (the vacuum in the Fock space). The condition that $P - \Pi \in \mathfrak{S}_2(\mathfrak{H})$ is usually called the "Shale-Stinespring" condition.

In this chapter we define the spaces $\mathfrak{S}_1^\Pi(\mathfrak{H})$ and the associated Π-trace. Then we give a decomposition for states Q belonging to $\mathcal{K}$. In particular we present a useful formula when $Q = P - \Pi$ with $P^2 = P$. Finally we recall the definition and properties of the Fock space associated with Π and we link our results with previously known statements in the Fock representation. The presented results are all contained (with their proof) in our works [6, 7]. They could be used as a basis for any quantum theory of a system whose Hartree-Fock (ground) state is a perturbation of a fixed reference HF state containing infinitely many particles.

A.1 Generalization of the trace functional

A.1.1 Definition of Π-trace

Definition A.1. *Let* Π *be a projector on* $\mathfrak{H}$ *such that* Π *and* $1-\Pi$ *have infinite rank, and* $A \in \mathfrak{S}_2(\mathfrak{H})$. *We shall say that* A *is* Π-trace class *if and only if* $A^{++} := (1-\Pi)A(1-\Pi)$ *and* $A^{--} := \Pi A \Pi$ *are trace class. Then we define the* Π-trace of A *by*

$$\mathrm{Tr}_\Pi(A) := \mathrm{Tr}(A^{++}) + \mathrm{Tr}(A^{--}).$$

We denote by $\mathfrak{S}_1^\Pi(\mathfrak{H})$ *the space of all Hilbert-Schmidt operators which are* Π*-trace class.*

Notice that if A is a trace class operator, then $A \in \mathfrak{S}_1^\Pi(\mathfrak{H})$ and $\mathrm{Tr}(A) = \mathrm{Tr}_\Pi(A)$ for any projector Π. However it is easy to see that a Π-trace class operator is not necessarily trace-class. Indeed we have seen in Chapter 4 and 5 that in general ground states of infinite quantum systems in a Hartree-Fock-type approximation are usually not trace-class.

Remark A.1. *We will see in Section A.2 that for* $Q \in \mathcal{K}$, $\mathrm{Tr}_\Pi(Q)$ *should be interpreted (up to a multiplication by* $-e$*) as the charge of the corresponding quantum state (counted relatively to that of* Π*). It has no sign* a priori.

Remark A.2. *It is obvious that* $\mathfrak{S}_1^\Pi(\mathfrak{H})$ *is a Banach space when endowed with the norm*

$$\|A\|_{\mathfrak{S}_1^\Pi(\mathfrak{H})} = \|A\|_{\mathfrak{S}_2(\mathfrak{H})} + \left\|A^{++}\right\|_{\mathfrak{S}_1(\mathfrak{H})} + \left\|A^{--}\right\|_{\mathfrak{S}_1(\mathfrak{H})}.$$

As $\mathfrak{S}_1(\mathfrak{H})$ *is the dual of the space of compact operators acting on* $\mathfrak{H}$, $\mathfrak{S}_1(\mathfrak{H})$ *can be endowed with the associated weak-*$*$ *topology where* $A_n \rightharpoonup A$ *in* $\mathfrak{S}_1(\mathfrak{H})$ *means that* $\mathrm{Tr}(A_nK) \to \mathrm{Tr}(AK)$ *for any compact operator* K. *Together with the fact that* $\mathfrak{S}_2(\mathfrak{H})$ *is a Hilbert space, this defines a weak topology on* $\mathfrak{S}_1^\Pi(\mathfrak{H})$ *for which any bounded sequence in* $\mathfrak{S}_1^\Pi(\mathfrak{H})$ *has a* $*$*-weak converging subsequence. It is clear that the convex set* $\mathcal{K}$ *defined in* (A.1) *is closed both for the strong and the* $*$*-weak topology of* $\mathfrak{S}_1^\Pi(\mathfrak{H})$. *However, the functional* $A \mapsto \mathrm{Tr}_\Pi(A)$ *is not weakly lower semi-continuous for this topology. Hence the sets (consisting of states having a fixed charge* $-eq$*)*

$$\mathcal{K}(q) = \{Q \in \mathcal{K} \mid \mathrm{Tr}_\Pi(Q) = q\} \tag{A.2}$$

are not closed for the $*$*-weak topology. Indeed for any* $q' \neq q$ *it is easy to construct a sequence* $Q_n \rightharpoonup Q$ *such that* $Q_n \in \mathcal{K}(q)$ *for all* n *and* $Q \in \mathcal{K}(q')$.

The following result shows that $\mathfrak{S}_1^\Pi(\mathfrak{H}) = \mathfrak{S}_1^{\Pi'}(\mathfrak{H})$ when $\Pi - \Pi' \in \mathfrak{S}_2(\mathfrak{H})$:

Lemma A.1 (Changing the reference projector [6])**.** *Let* Π *and* Π' *be two orthogonal projectors such that* $\Pi - \Pi' \in \mathfrak{S}_2(\mathfrak{H})$. *Then* A *is* Π*-trace class if and only if it is* Π'*-trace class, and in this case* $\mathrm{Tr}_\Pi(A) = \mathrm{Tr}_{\Pi'}(A)$.

Another important fact is that when A is Hilbert-Schmidt and $A + \Pi$ is a projector, then A has a Π-trace, as explained below:

Lemma A.2 (Generalized trace of a difference of two projectors [6])**.** *Let* P *and* Π *be two projectors on a Hilbert space* $\mathfrak{H}$, *such that* $P - \Pi$ *is a Hilbert-Schmidt operator. Then* $P - \Pi$ *is* Π*-trace class. Moreover,* $\mathrm{Tr}_\Pi(P - \Pi)$ *is an integer which satisfies*

$$\mathrm{Tr}_\Pi(P - \Pi) = \mathrm{Tr}\left((P - \Pi)^{2n+1}\right)$$

for all $n \geq 1$, *and* $\mathrm{Tr}_\Pi(P - \Pi) = 0$ *when* $\|P - \Pi\| < 1$.

The proof of the two lemmas may be found in [6, p. 528]. For Lemma A.2, we proved first that $\mathrm{Tr}_\Pi(P-\Pi) = \mathrm{Tr}(P-\Pi)^3$ and then used a result of Seiler and Simon [1] on pairs of projectors, giving the equality for every n. Indeed, a pair (P, Π) with $P - \Pi \in \mathfrak{S}_2(\mathfrak{H})$ is Fredholm, as defined in [1]. Lemma A.2 is also a consequence of Theorem A.1 that will be stated below.

The condition that $P - \Pi \in \mathfrak{S}_2(\mathfrak{H})$ is linked to the Shale-Stinespring theorem [15, 17], as we will explain in details in the next section. This means that the Fock representation associated with P is equivalent to that of Π. But before entering in these details, we quote some results of [7] which help in understanding better the geometry of $\mathfrak{S}_1^\Pi(\mathfrak{H})$.

A.1.2 A decomposition for $P - \Pi$

We start with a difference of two projectors. The decomposition below is valid in a more general setting (for any Fredholm pair of projections (P, Π) [1]) but for the sake of simplicity, we restrict ourselves to the Hilbert-Schmidt case needed in the text. We will deduce a general structure result for the variational set $\mathcal{K}$ in the next subsection.

Theorem A.1 (Reduction for difference of projections in $\mathcal{K}$ [7, Thm 5])**.** *Let P be an orthogonal projector on $\mathfrak{H}$ such that $P - \Pi \in \mathfrak{S}_2(\mathfrak{H})$. Denote by $(f_1, ..., f_N) \in (\mathfrak{H}_+)^N$ an orthonormal basis of $E_1 = \ker(P - \Pi - 1) = \ker(\Pi) \cap \ker(1 - P)$ and by $(g_1, ..., g_M) \in (\mathfrak{H}_-)^M$ an orthonormal basis of $E_{-1} = \ker(P - \Pi + 1) = \ker(1 - \Pi) \cap \ker(P)$. Then there exist an orthonormal basis $(v_i)_{i\geq 1} \subset \mathfrak{H}_+$ of $(E_1)^\perp$ in $\mathfrak{H}_+$, an orthonormal basis $(u_i)_{i\geq 1} \subset \mathfrak{H}_-$ of $(E_{-1})^\perp$ in $\mathfrak{H}_-$, and a sequence $(\lambda_i)_{i\geq 1} \in \ell_2(\mathbb{R}^+)$ such that*

$$P = \sum_{n=1}^N |f_n\rangle\langle f_n| + \sum_{i=1}^\infty \frac{|u_i + \lambda_i v_i\rangle\langle u_i + \lambda_i v_i|}{1+\lambda_i^2}, \tag{A.3}$$

$$1 - P = \sum_{m=1}^M |g_m\rangle\langle g_m| + \sum_{i=1}^\infty \frac{|v_i - \lambda_i u_i\rangle\langle v_i - \lambda_i u_i|}{1+\lambda_i^2}. \tag{A.4}$$

The numbers $\theta_i = \arccos(1+\lambda_i^2)^{-1/2}$ are usually called the Bogoliubov angles. The proof of Theorem A.1 was given in [7]. It is an adaptation of ideas in [1, 11, 17, 9, 8, 12]. Since $\Pi = \sum_{m=1}^M |g_m\rangle\langle g_m| + \sum_{i\geq 1} |u_i\rangle\langle u_i|$, we obtain

$$P - \Pi = \sum_{n=1}^N |f_n\rangle\langle f_n| - \sum_{m=1}^M |g_m\rangle\langle g_m| + \sum_{i\geq 1} \frac{\lambda_i^2}{1+\lambda_i^2}\big(|v_i\rangle\langle v_i| - |u_i\rangle\langle u_i|\big) \tag{A.5}$$

$$+ \sum_{i\geq 1} \frac{\lambda_i}{1+\lambda_i^2}\big(|u_i\rangle\langle v_i| + |v_i\rangle\langle u_i|\big). \tag{A.6}$$

The terms in (A.5) form the diagonal part of $P - \Pi$, which is trace-class by Lemma A.2. The last term (A.6) is the off-diagonal term which is only Hilbert-Schmidt *a priori.* Note we obtain from this formula that

$$\boxed{\mathrm{Tr}_\Pi(P - \Pi) = N - M}$$

is an integer as claimed in Lemma A.2. The formula of $P - \Pi$ can also be written as

$$P = \Pi + \sum_{n=1}^N |f_n\rangle\langle f_n| - \sum_{m=1}^M |g_m\rangle\langle g_m| + Q(A)$$

where $A := \sum_{i\geq 1} \lambda_i |v_i\rangle\langle u_i|$ and

$$Q(A) = \frac{A^*A}{1+A^*A} - \frac{AA^*}{1+AA^*} + A\frac{1}{1+A^*A} + \frac{1}{1+A^*A}A^*.$$

Therefore

$$\big\{P \mid P = P^* = P^2,\ P - \Pi \in \mathfrak{S}_2(\mathfrak{H}),\ \|P - \Pi\| < 1\big\} = \{\Pi + Q(A),\ A \in \mathfrak{S}_2(\mathfrak{H}_+, \mathfrak{H}_-)\}\,.$$

Remark A.3. *It is an easy consequence of Theorem A.1 that the connex components of the set of pure states*

$$\{P - \Pi \in \mathcal{Q} \mid P^2 = P\}$$

are indexed by their charge $\mathrm{Tr}_\Pi(P - \Pi) = N - M$. *This was already proved before in [14, 4].*

A.1.3 The structure of $\mathcal{K}$

We can now clarify the structure of the variational set $\mathcal{K}$ defined in (A.1).

Theorem A.2 (Structure of the variational set $\mathcal{K}$ [7, Thm 6])**.** *The set* $\mathcal{K}$ *coincides with the set containing all the operators of the form*

$$Q = U_D(\Pi + \gamma)U_{-D} - \Pi \tag{A.7}$$

where

1. $D \in \mathfrak{S}_2(\mathfrak{H})$ *is such that* $\ker D \supseteq \ker \Pi$ *and* $\ker D^* \supseteq \ker(1 - \Pi)$*;*
2. $U_D = \exp(D - D^*)$*;*
3. $\gamma \in \mathfrak{S}_1(\mathfrak{H})$ *is a self-adjoint and trace-class operator such that* $[\gamma, \Pi] = 0$ *and, denoting* $\gamma^{--} = \Pi\gamma\Pi$ *and* $\gamma^{++} = (1 - \Pi)\gamma(1 - \Pi)$*, then* $-\Pi \leq \gamma^{--} \leq 0$ *and* $0 \leq \gamma^{++} \leq 1 - \Pi$.

This result was used several times in our work to construct suitable trial states. A particularly important consequence is the following

Corollary A.1 (Approximation by finite-rank operators [7, Prop. 2])**.** *The set consisting of the operators* Q *which satisfy*

1. $Q \in \mathcal{K}(q)$*;*
2. $Q = P - \Pi + \gamma$ *where* P *is an orthogonal projector and* γ *is a finite rank operator such that* $0 \leq \gamma < 1$*,* $P\gamma = \gamma P = 0$*;*
3. Q *has a finite rank;*

is a dense subset of $\mathcal{K}(q)$ *for the strong topology of* $\mathfrak{S}_1^\Pi(\mathfrak{H})$.

We recall for the reader's convenience that $\mathcal{K}(q)$ was defined earlier in (A.2).

A.2 Bogoliubov states in the Fock space with vacuum Π

In this section we recall the definition of Bogoliubov states and we exhibit a link between the results of the previous section and known results in the Fock space.

The projector Π being given and fixed, we may define the following two subspaces of $\mathfrak{H}$:

$$\mathfrak{H}^+ := (1 - \Pi)\mathfrak{H} \quad \text{and} \quad \mathfrak{H}^- := \Pi\mathfrak{H}.$$

This allows us to introduce the space of N particles and the space of M holes as follows:

$$\mathcal{F}_+^N := \bigwedge_1^N \mathfrak{H}^+, \qquad \mathcal{F}_-^M := \bigwedge_1^M \mathcal{J}\mathfrak{H}^-$$

with the convention that $\mathcal{F}^0_\pm = \mathbb{C}$, and where $\mathcal{J}$ is any anti-linear unitary operator.[1] Eventually the Fock space associated with Π is defined as

$$\mathcal{F} := \bigoplus_{N,M\geq 0} \mathcal{F}^N_+ \otimes \mathcal{F}^M_-.$$

Acting on this space we have two creation and annihilation operators $b^\dagger_\pm(f)$ and $b_\pm(f)$ acting on the corresponding components of the Fock space. We use the following notation: $a^\dagger_+(f) = b^\dagger_+((1-\Pi)f)$ and $a^\dagger_-(f) = b^\dagger_-(\mathcal{J}\Pi f)$ for any $f \in \mathfrak{H}$. Note that $a^\dagger_+(\cdot)$ is linear whereas $a^\dagger_-(\cdot)$ is anti-linear.

Now we define the usual particle/anti-particle annihilation field operator as follows:

$$\Psi(f) := a_+(f) + a^\dagger_-(f)$$

(this operator is anti-linear). The vacuum $\Omega = 1 \in \mathcal{F}^0_+ \otimes \mathcal{F}^0_- = \mathbb{C}$ is characterized (up to a phase) by the properties $a_+(f)\Omega = a_-(f)\Omega = 0$ for any $f \in \mathfrak{H}$, and $\|\Omega\|_{\mathcal{F}} = 1$. Written in terms of Ψ, this becomes $\Psi((1-\Pi)f)\Omega = \Psi^\dagger(\Pi f)\Omega = 0$, for all $f \in \mathfrak{H}$.

A.2.1 The Shale-Stinespring condition

Let us now fix an orthogonal projector P acting on $\mathfrak{H}$. The natural question arises whether there exists a state $\Omega_P \in \mathcal{F}$ characterized by the relations

$$\Psi((1-P)f)\Omega_P = \Psi^\dagger(Pf)\Omega_P = 0 \quad \text{for all} \quad f \in \mathfrak{H}. \tag{A.8}$$

The answer is given by the celebrated Shale-Stinespring theorem [15]

Theorem A.3 (Shale-Stinespring [15]). *Let P be an orthogonal projector acting on $\mathfrak{H}$. There exists a state $\Omega_P \in \mathcal{F}$ such that* (A.8) *holds true, if and only if $P - \Pi \in \mathfrak{S}_2(\mathfrak{H})$. Then Ω_P is unique up to a phase. It may be written, using the notation of Theorem A.1,*

$$\begin{aligned}\Omega_P &= k \prod_{n=1}^N a^\dagger_+(f_n) \prod_{m=1}^M a^\dagger_-(g_m) \exp(A a^\dagger_+ a^\dagger_-)\,\Omega && \text{(A.9)}\\ &= k \prod_{n=1}^N a^\dagger_+(f_n) \prod_{m=1}^M a^\dagger_-(g_m) \prod_{i\geq 1} \big(1 + \lambda_i a^\dagger_+(v_i) a^\dagger_-(u_i)\big)\,\Omega && \text{(A.10)}\end{aligned}$$

where $A := \sum_{i\geq 1} \lambda_i |v_i\rangle\langle u_i|$ and $k = \prod_{i\geq 1}(1+\lambda_i^2)^{-1/2}$.

Formula (A.9) is classical and can be found in different forms in [17, 11, 12, 13, 14, 5] (see also [3, Theorem 2.2]). However we are not aware of any interpretation in terms of the projectors themselves, as expressed by Theorem A.1.

The unitary operator

$$\mathbb{U} := k \prod_{n=1}^N a^\dagger_+(f_n) \prod_{m=1}^M a^\dagger_-(g_m) \exp(A a^\dagger_+ a^\dagger_-)$$

is called a *Bogoliubov* transformation. It is not uniquely defined (one may add in the exponential terms of the form $a^\dagger_+ a_-$, $a_+ a^\dagger_-$ and $a_+ a_-$ without changing Ω_P).

---

[1] In the Dirac case, one usually chooses the charge conjugation operator. In general one can choose any anti-linear unitary operator, the most natural one being probably the dual map $\mathcal{J} : \mathfrak{H} \to \mathfrak{H}^*$ defined by $\mathcal{J}(f)(g) = \langle f, g\rangle$, as done in [16] and already mentioned in Chapter 2. If $\mathfrak{H} = L^2(\mathbb{R}^d, \mathbb{C}^q)$, $\mathcal{J}$ is just the usual (transposition and) complex conjugation of vector-valued functions.

A.2.2 Density matrices

In this section, we recall the definition of density matrices for any state ω acting on $\mathcal{F}$. This means ω is a linear form acting on the space of bounded operators $\mathcal{B}(\mathcal{F})$ such that $\omega(1) = 1$ and $\omega(B^*B) \geq 0$ for every B. We associate with ω two operators $\gamma : \mathfrak{H} \to \mathfrak{H}$ and $p : \mathcal{J}\mathfrak{H} \to \mathfrak{H}$ defined as follows:

$$\langle f, \gamma g\rangle = \omega\big(\Psi^\dagger(g)\Psi(f)\big), \qquad \langle f, p\mathcal{J}g\rangle = \omega\big(\Psi(g)\Psi(f)\big).$$

The operator γ is usually called the *one-body density matrix* whereas p is usually called the *pairing density matrix* of ω. It may be verified that

$$0 \leq \begin{pmatrix} \gamma & p \\ p^* & 1 - \mathcal{J}\gamma\mathcal{J}^* \end{pmatrix} \leq 1 \tag{A.11}$$

as an operator acting on $\mathfrak{H} \oplus \mathcal{J}\mathfrak{H}$. In particular we obtain that $0 \leq \gamma = \gamma^* \leq 1$. For any such state ω we may also introduce its density matrix counted with respect to the reference state Π defined as

$$Q := \gamma - \Pi$$

and which satisfies $-\Pi \leq Q \leq 1 - \Pi$, hence $Q^{++} \geq 0$, $Q - - \leq 0$.

It can be verified using Formulas (A.10), (A.3) and (A.4), that the density matrices of a Bogoliubov state Ω_P as defined in Theorem A.3 are respectively $\gamma = P$ and $p = 0$, as expected.

A.2.3 Charge in Fock space

As explained in the previous section, the density matrix of a Bogoliubov state of the form $\Omega_P = a_-^\dagger(g_1)\Omega$ is just $P = \Pi - |g_1\rangle\langle g_1|$. It really consists in removing a particle from the reference state Π or saying differently to create a hole in it. On the contrary the density matrix of $\Omega_P = a_+^\dagger(f_1)\Omega$ is $P = \Pi + |f_1\rangle\langle f_1|$, i.e. it consists in creating a particle.

Motivated by this observation, we may define the following operator

$$\mathcal{Q} := \sum_{i\geq 1} a_+^\dagger(e_i^+)a_+(e_i^+) - \sum_{i\geq 1} a_-^\dagger(e_i^-)a_-(e_i^-)$$

where $\{e_i^\pm\}$ is any orthonormal basis of $\mathfrak{H}^\pm$. This operator can be interpreted as the charge (up to a factor $-e$) counted relatively to the charge of Π. The minus in front of the second term accounts from the fact that creating a state in $\mathfrak{H}^-$ is the same as making a hole in the reference state, as mentioned before. Discarding the constant $-e$, $\mathcal{Q}$ is usually called the *charge operator*. Note that the operator $\mathcal{N} := |\mathcal{Q}|$ is just the usual particle number operator.

It can be checked that any Bogoliubov state Ω_P with $P - \Pi \in \mathcal{K}$ is in the form domain of $\mathcal{Q}$, i.e.

$$\langle \Omega_P, |\mathcal{Q}|\Omega_P\rangle = N + M + 2\sum_{i\geq 1} \frac{\lambda_i^2}{1 + \lambda_i^2} < \infty$$

and that

$$\boxed{\langle \Omega_P, \mathcal{Q}\Omega_P\rangle = \mathrm{Tr}_\Pi(P - \Pi) = N - M}$$

as expected. In particular, the average charge of a Bogoliubov state is always an integer.

A.2.4 Quasi-free states

In the previous subsections we have only considered pure Hartree-Fock states in the Fock space. We may indeed consider generalized states like in [3] and in Chapter 2. Like for the Bogoliubov states Ω_P, these states are completely characterized by their one-body density matrices.

Let ω be any state with one-body density matrix γ and renormalized density matrix $Q = \gamma - \Pi$. A simple calculation using the definition of $\mathcal{Q}$ shows that

$$\boxed{\omega(\mathcal{Q}) = \mathrm{Tr}_\Pi(Q) \quad \text{and} \quad \omega(|\mathcal{Q}|) = \mathrm{Tr}(Q^{++} - Q^{--})}.$$

Hence we have that $\omega(|\mathcal{Q}|) < \infty$ if and only if $Q \in \mathfrak{S}_1^\Pi(\mathfrak{H})$. For any such state ω with finite average number of particles, we therefore have $Q \in \mathcal{K}$.

A *quasi-free state* (with finite average particle number) [3, 2] is a state ω which satisfies the following two conditions:

1. ω is in the domain of $\mathcal{Q}$, i.e. $\omega(|\mathcal{Q}|) < \infty$;
2. for any $\{d_1, ..., d_{2K+1}\} \subset \{a_\pm(e_i^\pm)\} \cup \{a_\pm^\dagger(e_i^\pm)\}$ we have
$$\omega(d_1 \cdots d_{2K+1}) = 0$$
$$\omega(d_1 \cdots d_{2K}) = \sum_{\sigma \in S} \epsilon(\sigma)\omega(d_{\sigma(1)}d_{\sigma(2)}) \cdots \omega(d_{\sigma(2K-1)}d_{\sigma(2K)})$$
where S is the set of permutations such that $\sigma(1) < \sigma(3) < \cdots \sigma(2K-1)$ and $\sigma(2i-1) < \sigma(2i)$.

Here we have fixed as before a basis $\{e_i^\pm\}$ of $\mathfrak{H}^\pm$ but it can be checked that the previous definition is indeed independent of that basis. It was proved in [2, 3] that any such state is completely characterized by its two matrices γ and p.

Theorem A.4 (Quasi-free states [2, 3])**.** *For any $Q \in \mathcal{K}$ and any antisymmetric $p \in \mathfrak{S}_2(\mathcal{J}\mathfrak{H}, \mathfrak{H})$ such that* (A.11) *holds true, there exists a unique quasi-free state ω with finite particle number, $\omega(|\mathcal{Q}|) < \infty$, whose density matrix is $\Pi + Q$ and pairing density matrix is p.*

Hence restricting to quasi-free states amounts to working on the density matrices only. Usually for repulsive systems the pairing density matrix always vanishes for a minimizer [3], in which case one can restrict to states having no pairing. However, for (partially) attractive potentials, the pairing density matrix may be nonzero for a ground state.

A.2.5 Normal ordering

Normal ordering is an operation denoted as $: \, :_\Pi$ which consists in placing the creation operators on the left by anti-commuting them with the corresponding annihilation operator. For instance we have

$$: \Psi^\dagger(f)\Psi(g) :_\Pi = a_+^\dagger(f)a_+(g) + a_-(f)a_+(g) + a_+^\dagger(f)a_-^\dagger(g) - a_-^\dagger(g)a_-(f)$$

Note this uses the fact that $a_\pm^\dagger$ are operators of the Fock space built on Π, hence normal ordering depends on Π, a fact that we have strengthened in the notation. Now let us fix as before any orthogonal basis $\{e_i\}$ of $\mathfrak{H}$ and assume that each e_i belongs to $\mathfrak{H}^+$ or $\mathfrak{H}^-$. Then we can write

$$\mathcal{Q} = \sum_i : \Psi^\dagger(e_i)\Psi(e_i) :_\Pi .$$

In Physics one would use the notation

$$\mathcal{Q} = \int : \Psi^\dagger(x)\Psi(x) :_\Pi dx.$$

Similarly, it can be checked that

$$\langle e_j, Qe_i \rangle = \omega\left(: \Psi^\dagger(e_i)\Psi(e_i) :_\Pi\right) \tag{A.12}$$

for all states ω and any such basis $\{e_i\}$, where $Q = \gamma - \Pi$ is the one-body density matrix with Π as reference state.

Let us now consider another reference state Π' such that $\Pi' - \Pi \in \mathfrak{S}_2(\mathfrak{H})$, with associated state $\Omega_{\Pi'} \in \mathcal{F}$ in the Fock space. Due to the characterization of $\Omega_{\Pi'}$, we may define the associated creation and annihilation operators as

$$a^\dagger_{\Pi',+}(f) := \Psi^\dagger((1-\Pi')f) \quad \text{ and } \quad a^\dagger_{\Pi',-}(f) := \Psi(\Pi' f).$$

Note that $a^\dagger_{\Pi',+}$ has the correct linear behavior since $\Psi(\cdot)$ has been made linear by the introduction of the anti-linear operator $\mathcal{J}$. On the other hand, like for $a^\dagger_-$, the operator $a^\dagger_{\Pi',-}$ is anti-linear. Note we have as before

$$\Psi^\dagger(f) = a^\dagger_{\Pi',+}(f) + a_{\Pi',-}(f).$$

Hence we may define normal-ordering with respect to Π' as for Π and we obtain:

$$: \Psi^\dagger(f)\Psi(g) :_{\Pi'} = a^\dagger_{\Pi',+}(f)a_{\Pi',+}(g) + a_{\Pi',-}(f)a_{\Pi',+}(g) + a^\dagger_{\Pi',+}(f)a^\dagger_{\Pi',-}(g) - a^\dagger_{\Pi',-}(g)a_{\Pi',-}(f).$$

This leads to the following definition:

$$\mathcal{Q}_{\Pi'} := \sum_i : \Psi^\dagger(f_i)\Psi(f_i) :_{\Pi'}$$

for any basis $\{f_i\}$ composed of elements of $\Pi'\mathfrak{H}$ and $(1-\Pi')\mathfrak{H}$. Similarly to (A.12) we have $\langle f_j, Q' e_f\rangle = \omega\big(: \Psi^\dagger(f_i)\Psi(f_i) :_\Pi\big)$ where $Q' = \gamma - \Pi'$ is the density matrix of ω with reference Π'. We get for any state

$$\omega(\mathcal{Q}_{\Pi'}) = \mathrm{Tr}_{\Pi'}(Q') = \mathrm{Tr}_\Pi(Q') = \mathrm{Tr}_\Pi(Q + \Pi' - \Pi) = \omega(\mathcal{Q}) - \langle \Omega_{\Pi'}, \mathcal{Q}\Omega_{\Pi'}\rangle.$$

Hence we obtain that the charge behaves linearly under changes of the reference state

$$\mathcal{Q}_{\Pi'} = \mathcal{Q} - \langle \Omega_{\Pi'}, \mathcal{Q}\Omega_{\Pi'}\rangle$$

where we recall that the last term is an integer. This formula can easily be recovered using the canonical anti-commutation relations.

Let us now consider any bounded self-adjoint operator A acting on $\mathfrak{H}$ (we could also consider an unbounded operator but we restrict ourselves to the simplest case for convenience). We furthermore assume that $A^{+-} \in \mathfrak{S}_2(\mathfrak{H})$. We define the (normal-ordered) second quantization of A in $\mathcal{F}$ as

$$: A :_\Pi = \sum_{i,j} \langle e_i, A e_j\rangle : \Psi^\dagger(e_i)\Psi(e_j) :_\Pi .$$

As before, it can be verified using the definition of γ that for any state ω with finite average particle number, we have

$$\omega(: A :_\Pi) = \mathrm{Tr}_\Pi(AQ)$$

where $Q = \gamma - \Pi \in \mathcal{K}$. Using that A is bounded and that $A^{+-} \in \mathfrak{S}_2(\mathfrak{H})$ one verifies that this expression is well-defined for any state in the domain of $\mathcal{Q}$.

In principle the expectation value $\omega(: A :_\Pi)$ is not bounded from below when varying over all states, even when A is bounded. For instance for $A = 1$ we obtain $: 1 :_\Pi = \mathcal{Q}$ which is not bounded-below on $\mathcal{F}$. Indeed one can check that $\omega \mapsto \omega(: A :_\Pi)$ is bounded from below if and only if $\chi_{(-\infty,0)}(A) - \Pi \in \mathfrak{S}_2(\mathfrak{H})$, i.e. if and only if the Hartree-Fock state $P_A := \chi_{(-\infty,0)}(A)$ consisting of filling all the negative energies of A (which should clearly be the Hartree-Fock 'ground state') belongs to the same Fock space representation as our reference state Π. Assuming $P_A - \Pi \in \mathcal{K}$, we obtain for any state

$$\omega(: A :_\Pi) = \mathrm{Tr}_\Pi(AQ) = \mathrm{Tr}_{P_A}(AQ_A) + \mathrm{Tr}_\Pi(A(P_A - \Pi)) = \mathrm{Tr}_{P_A}(AQ_A) + \langle \Omega_{P_A}, : A :_\Pi \Omega_{P_A}\rangle$$

where we have introduced $Q_A := \gamma - P_A$. As it is clear that $\mathrm{Tr}_{P_A}(AQ_A) \geq 0$, we get that the above expression is minimized for the state Ω_{P_A}, as expected. When $\ker(A) = \{0\}$, it can even be shown that the minimizer is unique.

When $\chi_{(-\infty,0)}(A) - \Pi \notin \mathfrak{S}_2(\mathfrak{H})$, the previous argument can be adapted to get that the infimum of $\omega(: A :_\Pi)$ over all states is $-\infty$.

Typically we have a reference Hamiltonian A^0 which is perturbed by a potential V being A^0-compact. If one chooses as reference state $\Pi = \chi_{(-\infty,0]}(A^0)$, we will have that the ground state of the Hamiltonian $A^0 + V$ is formally $P_V := \chi_{(-\infty,0)}(A^0 + V)$. It lives in the same Fock representation as Π when $P_V - \Pi$ is Hilbert-Schmidt. In this case one can use the Fock space formalism and normal ordering to verify that P_V is really the ground state. For instance for $A^0 = D^0$, the Dirac operator, and V a real potential, necessary and sufficient conditions have been found by Klaus and Scharf [11] to ensure that $P_V - P_0 \in \mathfrak{S}_2(L^2(\mathbb{R}^3, \mathbb{C}^4))$. For the Coulomb potential (without an ultraviolet cut-off), it was proved [10] that this is not the case: $P_{\kappa|x|^{-1}} - P_0 \notin \mathfrak{S}_2(L^2(\mathbb{R}^3, \mathbb{C}^4))$.

As we have seen, our generalized trace functional mimics the properties of normal ordering in Fock space. Once a reference Hartree-Fock state has been chosen, this allows to give a mathematical meaning to certain physical quantities in the Fock space associated with our reference. This makes perfectly sense for *linear* minimization problems for which the ground state is always a Hartree-Fock state. However in general the true ground state of an *interacting* quantum system is never a Hartree-Fock state (it is not uniquely determined by its one-body density matrix), but instead a highly correlated state. In that case the Fock space formalism that we have presented cannot be used anymore and normal ordering is of no use.

References

[1] J. Avron, R. Seiler, and B. Simon, *The index of a pair of projections*, J. Funct. Anal., 120 (1994), pp. 220–237.

[2] V. Bach, J. M. Barbaroux, B. Helffer, and H. Siedentop, *On the stability of the relativistic electron-positron field*, Commun. Math. Phys., 201 (1999), pp. 445–460.

[3] V. Bach, E. H. Lieb, and J. P. Solovej, *Generalized Hartree-Fock theory and the Hubbard model*, J. Statist. Phys., 76 (1994), pp. 3–89.

[4] A. Carey and D. O'Brien, *Absence of vacuum polarisation in Fock space.*, Lett. Math. Phys., 6 (1982), pp. 339–340.

[5] H. Fierz and G. Scharf, *Particle interpretation for external field problems in QED*, Helv. Phys. Acta, 52 (1979), pp. 437–453 (1980).

[6] C. Hainzl, M. Lewin, and É. Séré, *Existence of a stable polarized vacuum in the Bogoliubov-Dirac-Fock approximation*, Commun. Math. Phys., 257 (2005), pp. 515–562.

[7] ———, *Existence of atoms and molecules in the mean-field approximation of no-photon quantum electrodynamics*, Arch. Rational Mech. Anal., 192 (2009), pp. 453–499.

[8] T. Kato, *Notes on projections and perturbation theory*, Tech. Rep. 9, University of California, 1955.

[9] ———, *Perturbation theory for linear operators*, Springer, second ed., 1995.

[10] M. Klaus, *Nonregularity of the Coulomb potential in quantum electrodynamics*, Helv. Phys. Acta, 53 (1980), pp. 36–39.

[11] M. Klaus and G. Scharf, *The regular external field problem in quantum electrodynamics*, Helv Phys. Acta, 50 (1977), pp. 779–802.

[12] S. N. M. Ruijsenaars, *On Bogoliubov transformations for systems of relativistic charged particles*, J. Math. Phys., 18 (1977), pp. 517–526.

[13] G. Scharf and H. P. Seipp, *Charged vacuum, spontaneous positron production and all that*, Physics Letters B, 108 (1982), pp. 196–198.

[14] H. P. Seipp, *On the S-operator for the external field problem of QED*, Helv. Phys. Acta, 55 (1982/83), pp. 1–28.

[15] D. Shale and W. F. Stinespring, *Spinor representations of infinite orthogonal groups*, J. Math. Mech., 14 (1965), pp. 315–322.

[16] J. P. Solovej, *Many body quantum mechanics.* Lectures notes at LMU Münich, 2007.

[17] B. Thaller, *The Dirac equation*, Texts and Monographs in Physics, Springer-Verlag, Berlin, 1992.